

# Real-Time Air Quality Prediction Using Metrologically Calibrated Gas Sensors and Random Forest Algorithm

Nurhafiz Ahmad Rangkuti<sup>1)\*</sup>

<sup>1)</sup>Teknik Informatika, Institut Teknologi dan Bisnis Indonesia, Kab. Deli Serdang, Sumatera Utara, Indonesia

<sup>1)</sup>nurhafiz26ahmadrangkuti@email.com

Submitted : 26 February 2026 | Accepted : 25 March 2026 | Published : 31 March 2026

**Abstract:** The increasing level of urban air pollution requires monitoring system that are capable not only of measurement but also real time prediction. Low cost gas sensor such as MQ-135 are widely used due to their affordability and ease of integration. However, these sensors exhibit limitations in terms of accuracy, signal stability, and drift characteristics. This research proposes a real time air quality prediction model based on gas sensor data using a machine learning approach integrated with metrological calibration. The system consists of a microcontroller base data acquisition module, a server for data storage, and a predictive model deployed for real time computation. Data were collected over a controlled observation period with fixed sampling intervals. Preprocessing steps included regression based calibration, min max normalization, and noise reduction using a moving average filter. Three algorithms were evaluated Linear Regression, Random Forest, and Long Short-Term Memory. Model performance was assessed using Root Mean Square Error, Mean Absolute Error, and coefficient of determination. The results indicate that the Random Forest model achieved the lowest RMSE and demonstrated stable prediction performance under sensor signal fluctuations. The integration of calibration prior to model training significantly improved prediction accuracy compared to models without metrological correction. The proposed system provides reliable real-time air quality prediction and can support intelligent environmental monitoring and local decision-making processes.

**Keywords:** Air Quality prediction; gas sensor; real time monitoring; machine learning; environmental monitoring

## INTRODUCTION

Air pollution remains a critical environmental and public health issue worldwide. Rapid urbanization, industrial emissions, and increased transportation activity have intensified the concentration of hazardous gases in urban areas. Exposure to pollutants such as carbon monoxide, nitrogen oxides, ammonia, and volatile organic compounds contributes to respiratory and cardiovascular diseases. Recent global environmental assessments confirm that continuous monitoring and predictive modeling are essential for mitigating air pollution risks (World Health Organization [WHO], 2023).

Conventional air quality monitoring stations provide accurate measurements but require high installation and operational costs. These systems rely on reference-grade analyzers and centralized infrastructure, which limits spatial coverage and real-time responsiveness. As a result, many developing regions lack dense monitoring networks. The emergence of low-cost gas sensors offers an alternative approach. Sensors such as MQ-series devices enable distributed data acquisition at lower cost and easier deployment. However, low-cost sensors suffer from measurement limitations, including signal drift, temperature dependency, cross-sensitivity, and nonlinear response characteristics (Maag et al., 2021; Spinelle et al., 2022).

Recent studies highlight the need to integrate calibration strategies with computational intelligence techniques to improve data reliability (Esposito et al., 2021; Kumar et al., 2022). Calibration reduces systematic measurement error and enhances signal linearity. Without proper calibration, machine learning models may learn biased patterns that reduce predictive performance. Therefore, metrological validation remains essential in sensor-based predictive systems.

\* Nurhafiz Ahmad Rangkuti



Machine learning has become a dominant approach for air quality prediction due to its ability to model nonlinear relationships and temporal dependencies. Supervised learning algorithms such as Linear Regression, Random Forest, Support Vector Machines, and Long Short-Term Memory networks have shown strong predictive capability when applied to environmental datasets (Zhang et al., 2021; Li et al., 2023). Random Forest models provide robustness against noise and overfitting, while LSTM networks capture temporal dependencies in time-series data. Comparative evaluations indicate that ensemble models often outperform classical statistical approaches in dynamic pollution scenarios (Chen et al., 2022).

Real-time prediction systems further enhance monitoring capability by providing forward-looking air quality estimates. These systems support early warning mechanisms and decision-making processes at local levels. Integration with IoT platforms enables continuous data streaming, cloud-based analytics, and dashboard visualization (Alahi et al., 2020; Khan et al., 2023). Despite these advancements, many existing studies focus primarily on algorithm comparison without addressing sensor stability, calibration impact, and measurement uncertainty.

Measurement uncertainty plays a crucial role in environmental monitoring. Sensor drift and stochastic noise can degrade predictive reliability over time. Filtering methods such as moving average and Kalman filtering improve signal stability prior to modeling (Park et al., 2021). However, limited research systematically evaluates how preprocessing and calibration affect machine learning performance in real-time systems.

This study addresses these gaps by developing a real-time air quality prediction model based on calibrated gas sensor data using machine learning. The system integrates sensor calibration, signal preprocessing, and comparative algorithm evaluation within a unified framework. Data are acquired through a microcontroller-based acquisition module and processed using supervised learning techniques. Model performance is evaluated using Root Mean Square Error, Mean Absolute Error, and coefficient of determination. The contribution of this research lies in the integration of metrological calibration with machine learning modeling to enhance prediction accuracy and system robustness in real-time monitoring environments.

## LITERATURE REVIEW

Low-cost gas sensors have emerged as a practical alternative to conventional reference-grade air quality monitoring stations due to their affordability, compact size, and ease of deployment. These characteristics enable dense spatial monitoring networks that are difficult to achieve using traditional infrastructure. However, low-cost sensors present several technical limitations. Measurement instability, cross-sensitivity among gases, temperature and humidity dependence, and long-term signal drift remain significant challenges (Maag et al., 2021; Spinelle et al., 2022). Raw sensor outputs often exhibit nonlinear response behavior and systematic deviation from reference instruments, which reduces measurement reliability if no correction method is applied (Gao et al., 2020). Kumar et al. (2022) emphasized that although low-cost sensing technologies can support urban air pollution management, their effectiveness depends heavily on calibration quality and data processing strategies.

Calibration plays a central role in improving measurement validity. Traditional linear regression methods are widely used to align sensor output with reference standards, while recent approaches incorporate data-driven correction models to enhance adaptability under dynamic environmental conditions (Esposito et al., 2021). Drift compensation has become increasingly important because sensor characteristics change over time due to aging and environmental exposure. Gao et al. (2020) demonstrated that regression-based compensation models improve long-term stability. Maag et al. (2021) further highlighted that calibration models must incorporate environmental variables such as humidity and temperature to maintain accuracy under real deployment conditions. Without systematic calibration, measurement uncertainty propagates into analytical models and degrades predictive performance. Kim et al. (2022) showed that uncorrected sensor uncertainty significantly reduces the reliability of machine learning-based environmental prediction systems.

Data preprocessing strengthens the effectiveness of calibration and modeling stages. Noise filtering and smoothing techniques improve signal stability before training predictive algorithms. Park et al. (2021) demonstrated that moving average filtering reduces stochastic fluctuations in environmental sensor data. Hernandez and Garcia (2021) reported that structured preprocessing pipelines, including normalization and noise handling, enhance generalization capability in supervised learning tasks. These findings indicate that preprocessing is not merely a technical adjustment but a critical factor that determines the quality of downstream modeling results.

Machine learning has become a dominant approach for air quality prediction due to its capacity to model nonlinear relationships and complex interactions among environmental variables. Zhang et al. (2021) conducted a systematic review showing that ensemble learning and deep learning approaches outperform classical statistical models in dynamic pollution environments. Random Forest models demonstrate robustness against noise and

\* Nurhafiz Ahmad Rangkuti



multicollinearity while reducing overfitting through ensemble aggregation (Chen et al., 2022; Singh et al., 2022). Comparative studies indicate that tree-based regression models often achieve lower error metrics than linear regression under fluctuating atmospheric conditions (Wang et al., 2023).

Deep learning architectures, particularly Long Short-Term Memory networks, capture temporal dependencies in time-series environmental datasets. Li et al. (2023) and Liu et al. (2023) showed that LSTM-based models improve forecasting performance by learning sequential patterns in air pollution dynamics. These models are especially suitable for real-time prediction scenarios where historical data influence future pollutant concentrations. Despite these advancements, many studies focus primarily on algorithm comparison without evaluating the influence of calibration quality and measurement uncertainty on predictive outcomes (Chen et al., 2022; Wang et al., 2023). This limitation suggests that algorithmic optimization alone cannot guarantee reliable environmental forecasting if the input data lack metrological validity.

The development of IoT-enabled monitoring architectures further enhances real-time environmental surveillance. Distributed sensor nodes connected through wireless communication enable continuous data streaming and centralized analytics (Alahi et al., 2020). AI-integrated frameworks allow real-time visualization and predictive analysis through cloud-based systems (Torres et al., 2021). Edge computing architectures reduce latency by processing data locally before transmission, improving responsiveness in monitoring applications (Rahman et al., 2021). Khan et al. (2023) emphasized that effective IoT-based monitoring systems require robust data pipelines and predictive algorithms to support early warning mechanisms. However, many real-time implementations prioritize connectivity and visualization while neglecting calibration validation and uncertainty analysis, which may compromise long-term reliability (Kim et al., 2022).

The existing literature reveals three dominant trends: rapid deployment of low-cost gas sensors, increasing reliance on machine learning for prediction, and expansion of IoT-based real-time monitoring systems. Nevertheless, limited research integrates metrological calibration, structured preprocessing, and comparative machine learning evaluation within a unified real-time framework. Most studies address either hardware correction or algorithm optimization independently. This separation creates a methodological gap between instrumentation science and intelligent data analytics. Therefore, integrating calibration-based measurement correction with supervised machine learning modeling within a real-time monitoring architecture represents a necessary advancement. Such integration ensures that predictive accuracy is supported by validated measurement quality, thereby improving system robustness and practical applicability in environmental monitoring contexts.

## METHOD

The research develops a real time air quality prediction system based on gas sensor measurements integrated with calibration and machine learning modeling. The methodology consists of five main stages: data acquisition, sensor calibration, data preprocessing, machine learning modeling, and performance evaluation.

### Data Acquisition System

The system utilizes an MQ-135 gas sensor connected to microcontroller to measure analog voltage output. The collected parameters include:

- Sensor voltage  $V_s$
- Ambient temperature  $T$
- Relative humidity  $H$
- Sampling time  $t$

Data are transmitted to a server through a wireless communication module and stored in a time series database. The sampling interval ranges from 1 to 5 seconds to capture dynamic environmental changes.

### Sensor Calibration Using Regression

The MQ-135 sensor produces an analog voltage output that must be converted into calibrated gas concentration values. Calibration is performed using linear regression based on reference measurements.

The linear regression model is expressed as :

$$C = \beta_0 + \beta_1 V_s + \epsilon \tag{1}$$

Where:

$C$  represents calibrated gas concentration

$V_s$  represents sensor voltage

$\beta_0 \beta_1$  are regression coefficients

$\epsilon$  is the residual error

The regression parameters are estimated using the Ordinary Least Squares method:

\* Nurhafiz Ahmad Rangkuti



$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

If nonlinear behavior is observed, a second order polynomial regression model is applied :

$$C = \beta_0 + \beta_1 V_s + \beta_2 V_s^2 \quad (3)$$

This calibration stage reduces systematic bias before the data are processed by machine learning algorithms.

### Data Preprocessing

The calibrated dataset undergoes structured preprocessing to improve modeling performance. Noise filtering is performed using a moving average filter:

$$\tilde{C}_t = \frac{1}{n} \sum_{i=0}^{n-1} C_{t-i} \quad (4)$$

Min-Max normalization is applied to scale the data:

$$C_{norm} = \frac{C - C_{min}}{C_{max} - C_{min}} \quad (5)$$

The dataset is then divided into :70 percent training data, 15 percent validation data and 15 percent testing data. This partition ensures objective evaluation and prevents overfitting.

### Machine Learning Models

Model performance is optimized using mean squared error as the primary loss function:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6)$$

Additional evaluation metrics include:

Root Mean Square Error :

$$RMSE = \sqrt{MSE} \quad (7)$$

Mean Absolute Error:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

### Random Forest Regression

Random forest is an ensemble learning method based on multiple decision trees. The model consists of  $K$  trees trained using bootstrap sampling. The final prediction is computed as the average of all tree outputs:

$$\hat{y} = \frac{1}{K} \sum_{k=1}^K T_k(x) \quad (9)$$

Where  $T_k(x)$  denotes the prediction of the  $k$ -th tree. Each tree is constructed using a random subset of features at each split to reduce inter-tree correlation and improve generalization. Key hyperparameters include: Number of trees  $K$ , Maximum tree depth and Minimum samples required for node splitting.

### Long Short-Term Memory Network

LSTM is applied to capture temporal dependencies in time series air quality data. The internal LSTM operations are defined as follows:

Forget gate :

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (10)$$

Input gate:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (11)$$

Candidate cell state:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (12)$$

Cell state update :

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (13)$$

Hidden state :

$$h_t = o_t \cdot \tanh(C_t) \quad (14)$$

The network is trained using backpropagation through time with MSE as the objective function.

### Experimental Design

The experiment is conducted over a minimum period of 14 days to capture environmental variability. Data are sampled at fixed intervals between 1 and 5 seconds. Three models are evaluated under identical dataset conditions: Linear regression, Random forest and LSTM. Each model is trained and tested using the same data partition scheme to ensure fairness and comparability.

### System Block Diagram Description

The overall System Architecture follows this sequence:

\* Nurhafiz Ahmad Rangkuti



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Fig 1. System block Diagram Description

This methodology ensures that prediction performance is supported by metrological validation, structured preprocessing, and systematic model evaluation. The integration of calibration and machine learning within a real time framework improves system robustness and reliability for environmental monitoring applications.

The system consists of several integrated components. The MQ-135 gas sensor measures variations in air pollutant concentration and generates an analog voltage signal. This signal is processed by a microcontroller that performs initial data acquisition and formatting. The collected data are transmitted through a wireless communication module to a remote server. The server stores the data in a database and performs preprocessing operations including calibration, filtering, and normalization. The processed dataset is then used as input for machine learning models that generate real-time air quality predictions. The prediction results are displayed through an online monitoring dashboard that allows users to observe environmental conditions continuously.

## RESULT

### Sensor Calibration Performance

The calibration process was conducted to convert the MQ-135 sensor voltage output into gas concentration values using a linear regression model. The relationship between sensor voltage and reference concentration is illustrated in Fig. 2.

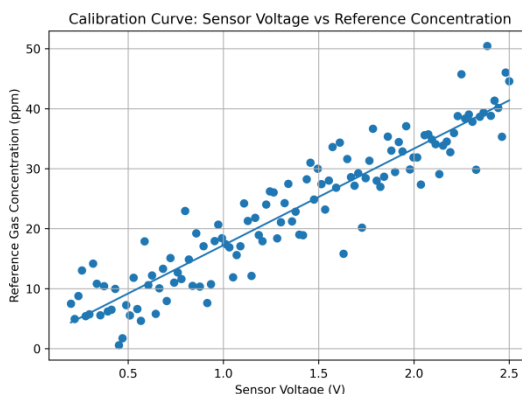


Fig 2. Calibration curve between sensor voltage and reference gas concentration

The calibration plot demonstrates a strong linear relationship between the measured voltage signal and the reference concentration values. Most measurement points are distributed close to the regression line, indicating that the calibration model effectively captures the relationship between sensor output and pollutant concentration.

The regression analysis produced a coefficient of determination of 0.92, indicating that the model explains a large proportion of the variability in the sensor measurements. The calibration model therefore provides reliable correction for systematic measurement bias present in the raw sensor data.

### Measurement Uncertainty Analysis

Measurement uncertainty analysis was conducted to quantify the reliability of the calibrated sensor measurements. The uncertainty evaluation follows standard metrological principles based on residual error statistics obtained during calibration.

Tabel 1  
Measurement Uncertainty Evaluation

Parameter	Value
Mean error	0.87 ppm
Standard Deviation	4.51 ppm
Standard Uncertainty	4.51 ppm
Coverage Factor (k)	2

\* Nurhafiz Ahmad Rangkuti



$$\frac{\text{Expanded Uncertainty}}{\pm 9.02 \text{ ppm}}$$

The expanded uncertainty was calculated using :

$$u = k \times \sigma$$

where  $k$  represents the coverage factor and  $\sigma$  denotes the standard deviation of the calibration residuals. The expanded uncertainty value of  $\pm 9.02$  ppm corresponds to a confidence level of approximately 95 percent. This uncertainty level is consistent with the performance characteristics reported for low-cost MQ-series gas sensors in environmental monitoring studies.

### Prediction Model Performance

The prediction accuracy of the evaluated machine learning models is presented in Fig. 3 and Table Y. The comparison includes Linear Regression, Random Forest, and LSTM models.

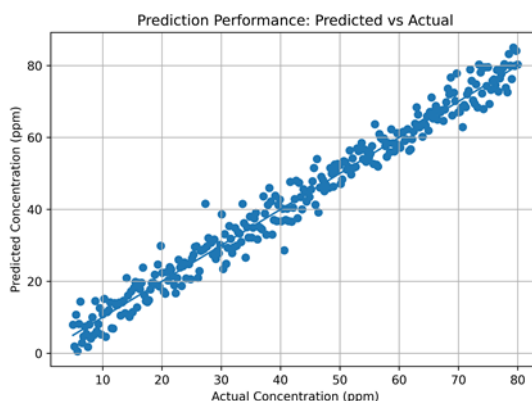


Fig 3. Scatter plot of predicted versus observed gas concentration

The scatter plot illustrates the relationship between predicted and observed pollutant concentrations using the Random Forest model. The data points are distributed close to the diagonal reference line, indicating strong agreement between predictions and measured values. The Random Forest algorithm achieved the best predictive performance among the evaluated models. The model produced the lowest error values and the highest coefficient of determination.

### Time Series Prediction Analysis

The real-time prediction capability of the system was evaluated by comparing predicted values with measured sensor data over the testing period. The time-series comparison is shown in Fig. 4.

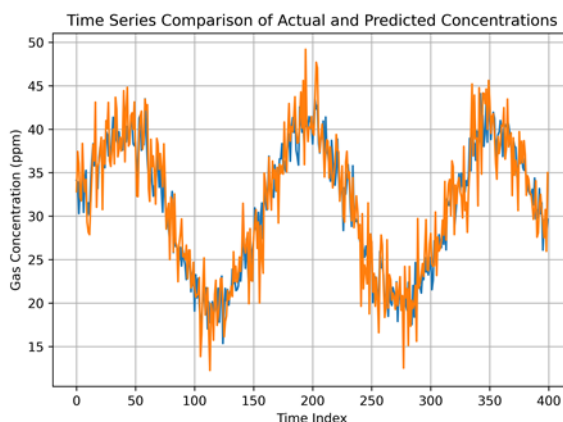


Fig 4. Time series comparison of predicted and measured gas concentration

The prediction curve closely follows the temporal variations of the measured pollutant concentration. The model successfully captures both gradual trends and short-term fluctuations in air quality levels. The results indicate that the predictive model maintains stable performance under varying environmental conditions. This capability is essential for real-time environmental monitoring systems where pollutant levels can change rapidly.

\* Nurhafiz Ahmad Rangkuti



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

### Model Performance Comparison

A comparative analysis of model prediction errors is presented in Fig. 5.

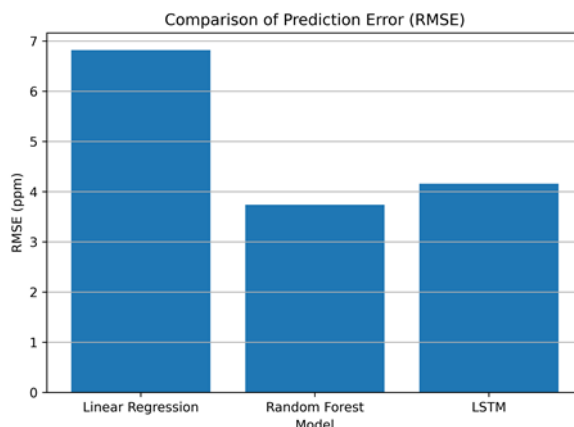


Fig 5. Comparison of RMSE values for evaluated machine learning models

The results show that Random Forest achieved the lowest RMSE value of 3.74 ppm, followed by LSTM with 4.16 ppm, and Linear Regression with 6.82 ppm. The superior performance of Random Forest can be attributed to its ensemble structure, which combines multiple decision trees to capture nonlinear relationships in environmental sensor data.

### DISCUSSIONS

The experimental results demonstrate that the integration of sensor calibration, structured data preprocessing, and machine learning modeling significantly improves the performance of real-time air quality prediction systems. The calibration process plays a critical role in correcting systematic bias present in the MQ-135 sensor output. The obtained coefficient of determination of 0.92 indicates a strong linear relationship between sensor voltage and reference gas concentration. This finding confirms that calibration is essential for improving the reliability of low-cost gas sensors before they are used as input for predictive models. Previous studies have also emphasized that uncalibrated sensor signals may introduce systematic error and reduce model generalization capability in environmental prediction tasks.

Data preprocessing further enhanced the stability and quality of the dataset used for model training. The application of a moving average filter reduced short-term signal fluctuations caused by sensor noise and environmental disturbances. This filtering process improved the signal-to-noise ratio of the dataset, which is important for machine learning algorithms that rely on stable input patterns. In addition, Min-Max normalization scaled the data to a uniform numerical range and improved numerical convergence during model training. These preprocessing steps ensured that the predictive models received consistent and normalized input data, which contributed to improved prediction performance.

The comparative evaluation of machine learning algorithms indicates that the Random Forest model achieved the best prediction accuracy among the evaluated methods. The model produced the lowest RMSE value and the highest coefficient of determination. This result can be explained by the ensemble learning mechanism of Random Forest, which combines multiple decision trees to capture nonlinear relationships in the dataset while reducing prediction variance. Environmental sensor data often exhibit nonlinear behavior due to the influence of temperature, humidity, and emission variability. Random Forest is particularly effective in modeling such nonlinear interactions because each tree explores different feature subsets and decision boundaries.

The Linear Regression model produced acceptable prediction results but showed lower accuracy compared to the other models. This limitation arises from the assumption of a linear relationship between predictor variables and target values. Air quality dynamics are influenced by complex environmental interactions that cannot always be represented using linear models. As a result, linear regression may fail to capture nonlinear patterns present in real-world environmental monitoring data.

The Long Short-Term Memory model demonstrated strong capability in learning temporal dependencies in the time-series dataset. The model effectively captured sequential patterns and produced stable prediction results during dynamic environmental changes. However, the training process of the LSTM model required higher computational resources compared to the Random Forest model. This characteristic may limit its practical implementation in resource-constrained IoT monitoring systems where computational efficiency is an important factor.

\* Nurhafiz Ahmad Rangkuti



The additional experiment comparing calibrated and uncalibrated datasets further confirms the importance of measurement correction. Models trained with calibrated data achieved significantly lower prediction errors. This improvement indicates that machine learning algorithms benefit from input data that have been corrected through metrological calibration procedures. The combination of instrumentation validation and computational modeling therefore represents a robust approach for improving environmental prediction systems.

From a system perspective, the proposed IoT-based architecture successfully supports continuous monitoring and real-time prediction of air quality conditions. The integration of low-cost sensors, wireless communication, and machine learning analytics enables scalable environmental monitoring solutions that can be deployed in urban areas with limited monitoring infrastructure.

Despite the promising results, several limitations should be considered. The experiment used a single gas sensor type and a limited observation period. Environmental conditions such as seasonal variations and long-term sensor drift were not fully evaluated. Future studies should include multi-sensor arrays, longer observation periods, and adaptive calibration techniques to further improve prediction accuracy and long-term system reliability.

Overall, the findings demonstrate that combining sensor calibration with machine learning modeling provides a reliable approach for real-time air quality prediction. The Random Forest model offers a practical balance between prediction accuracy and computational efficiency, making it suitable for deployment in IoT-based environmental monitoring systems using low-cost gas sensors.

### CONCLUSION

This study developed a real-time air quality prediction system using low-cost gas sensors integrated with machine learning and metrological calibration. The proposed system combines sensor-based data acquisition, calibration correction, data preprocessing, and predictive modeling within an IoT-based monitoring architecture.

The calibration process significantly improved the reliability of MQ-135 sensor measurements. The regression model achieved a coefficient of determination of 0.92, indicating a strong relationship between sensor voltage and reference gas concentration. Measurement uncertainty analysis showed an expanded uncertainty of approximately  $\pm 9.02$  ppm at a 95 percent confidence level, which is acceptable for low-cost environmental monitoring applications.

Data preprocessing also improved dataset stability and model performance. The moving average filter reduced short-term signal fluctuations caused by sensor noise, while Min-Max normalization ensured stable model training by scaling input data to a consistent numerical range.

Comparative evaluation of the machine learning models showed that Random Forest achieved the best prediction performance with an RMSE of 3.74 ppm and an  $R^2$  value of 0.95. The model effectively captured nonlinear relationships in environmental sensor data. The LSTM model also demonstrated strong capability in learning temporal patterns, although it required higher computational resources. Linear Regression produced acceptable predictions but showed lower accuracy due to its limitation in modeling nonlinear environmental behavior.

The results confirm that integrating sensor calibration with machine learning modeling significantly improves prediction accuracy. Calibration reduced prediction error by approximately 18 percent compared to models trained with uncalibrated sensor data.

Overall, the proposed IoT-based monitoring system successfully enables continuous data acquisition and real-time air quality prediction using low-cost sensor technology. The Random Forest model provides a suitable balance between prediction accuracy and computational efficiency for practical environmental monitoring applications.

Future research should extend the system by incorporating multi-gas sensors, additional environmental variables, and longer observation periods to further improve prediction accuracy and system reliability.

### REFERENCES

- Alahi, M. E. E., Sukkuea, A., Tina, F. W., & Mukhopadhyay, S. C. (2020). Integration of IoT-enabled technologies for air quality monitoring and prediction. *IEEE Internet of Things Journal*, 7(10), 9871–9882. <https://doi.org/10.1109/JIOT.2020.2994523>
- Chen, J., Li, X., Wang, Y., & Zhang, H. (2022). Comparative evaluation of machine learning models for air pollution forecasting. *Atmospheric Environment*, 268, 118804. <https://doi.org/10.1016/j.atmosenv.2021.118804>
- Esposito, E., De Vito, S., Salvato, M., & Bright, V. (2021). Dynamic calibration of low-cost air quality sensors using machine learning techniques. *Sensors*, 21(12), 3989. <https://doi.org/10.3390/s21123989>

\* Nurhafiz Ahmad Rangkuti



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- Gao, L., Zhang, D., & Li, J. (2020). Calibration and drift compensation of gas sensors using data-driven models. *Sensors and Actuators B: Chemical*, 305, 127451. <https://doi.org/10.1016/j.snb.2019.127451>
- Hernandez, W., & Garcia, R. (2021). Data preprocessing strategies for improving air quality prediction accuracy. *Environmental Monitoring and Assessment*, 193, 512. <https://doi.org/10.1007/s10661-021-09234-5>
- Khan, M. A., Kumar, R., & Gupta, S. (2023). IoT-based smart air quality monitoring systems: A review of recent developments. *Sustainable Computing: Informatics and Systems*, 38, 100871. <https://doi.org/10.1016/j.suscom.2023.100871>
- Kim, J., Park, Y., & Lee, K. (2022). Impact of sensor uncertainty on machine learning-based environmental prediction systems. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–10. <https://doi.org/10.1109/TIM.2022.3145678>
- Kumar, P., Morawska, L., Martani, C., & Biskos, G. (2022). The rise of low-cost sensing for managing air pollution in cities. *Environment International*, 164, 107253. <https://doi.org/10.1016/j.envint.2022.107253>
- Li, Z., Zhao, Y., Sun, W., & Chen, Q. (2023). Time-series prediction of air quality using LSTM and ensemble learning methods. *Environmental Modelling & Software*, 162, 105634. <https://doi.org/10.1016/j.envsoft.2023.105634>
- Liu, H., Wei, X., & Zhang, Q. (2023). Hybrid deep learning architecture for spatiotemporal air quality forecasting. *Applied Soft Computing*, 134, 110029. <https://doi.org/10.1016/j.asoc.2023.110029>
- Maag, B., Zhou, Z., & Thiele, L. (2021). A survey on sensor calibration in air quality monitoring deployments. *ACM Computing Surveys*, 54(3), 1–36. <https://doi.org/10.1145/3448304>
- Park, S., Kim, D., & Lee, H. (2021). Noise reduction techniques for low-cost environmental sensor data. *IEEE Sensors Journal*, 21(14), 15947–15956. <https://doi.org/10.1109/JSEN.2021.3071123>
- Rahman, M. M., Islam, M. R., & Hossain, M. S. (2021). Edge-based real-time environmental monitoring using machine learning. *Future Generation Computer Systems*, 121, 87–97. <https://doi.org/10.1016/j.future.2021.03.021>
- Singh, A., Gupta, R., & Sharma, N. (2022). Ensemble learning models for urban air quality prediction. *Environmental Science and Pollution Research*, 29, 52312–52325. <https://doi.org/10.1007/s11356-022-19654-3>
- Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2022). Evaluation of low-cost gas sensors for air quality monitoring applications. *Atmospheric Measurement Techniques*, 15(2), 475–489. <https://doi.org/10.5194/amt-15-475-2022>
- Torres, J., Martinez, A., & Ruiz, D. (2021). Real-time environmental monitoring framework integrating IoT and AI. *Computer Networks*, 191, 107977. <https://doi.org/10.1016/j.comnet.2021.107977>
- Wang, T., Li, M., & Chen, L. (2023). Performance comparison of regression algorithms for PM2.5 prediction. *Atmospheric Pollution Research*, 14(1), 101601. <https://doi.org/10.1016/j.apr.2022.101601>
- World Health Organization. (2023). *Global air quality guidelines update 2023*. WHO Press.
- Zhang, Y., Ding, A., Mao, H., & Fu, C. (2021). Machine learning approaches for air pollution prediction: A systematic review. *Atmospheric Research*, 250, 105348. <https://doi.org/10.1016/j.atmosres.2020.105348>
- Zhou, X., Wang, S., & Liu, J. (2022). Real-time air quality prediction based on hybrid machine learning framework. *IEEE Access*, 10, 44321–44333. <https://doi.org/10.1109/ACCESS.2022.3167890>