Jurnal Sistem Informasi Bisnis 04(2025)
 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/84892

Real-Time Air Quality Prediction Using IoT Sensor Networks and LSTM Deep Learning Model

Ahmad Heru Mujianto[a,*], Chamdan Mashuri[b]

[a b]Information Systems, Hasyim Asy'ari University, Jombang, East Java, Indonesia

Submitted: October 19th, 2025; Revised: December 9th, 2025; Accepted: December 12th, 2025; Available Online: December 29th, 2025

DOI: 10.21456/vol15iss4pp512-517

Abstract

The increasing number of motor vehicles in urban areas has contributed to declining air quality, affecting both public health and the environment. This condition becomes more critical at road intersections with high traffic density, particularly during morning rush hours. This study aims to develop a real-time air quality prediction system based on Internet of Things and the Long Short-Term Memory (LSTM) method. Air quality data were collected using IoT-based sensors installed at a road intersection with a traffic density of approximately 200 motor vehicles per minute between 06:30 and 07:30 AM. The observed parameters included Air Quality Index (AQI), temperature, humidity, and air pollutant concentrations. The research stages consisted of sensor data acquisition, data preprocessing using Min-Max normalization, time-series dataset construction using a sliding window approach, and LSTM model training for air quality forecasting. Experimental results showed that the LSTM model was capable of predicting air quality effectively based on temporal sensor data patterns. The evaluation results produced a Mean Absolute Error (MAE) value of 2.046 and a Root Mean Square Error (RMSE) value of 2.076. The findings demonstrate that the integration of IoT and LSTM is effective for real-time air quality monitoring and forecasting. The novelty of this study lies in the use of real-time sensor data collected from a high-traffic road intersection and the integration of monitoring and forecasting systems within a single deep learning-based platform. The proposed system has the potential to support smart environmental monitoring and early warning systems in urban areas.

Keywords : air quality, Internet of Things, LSTM, deep learning, real-time prediction

1. Introduction

Air quality is one of the important indicators in determining the level of environmental health and quality of life of the community. The increase in the number of motor vehicles in urban areas causes exhaust emissions to be higher and has a direct impact on air pollution. Pollutants such as carbon monoxide (CO), PM2.5 particulates, PM10, nitrogen dioxide (NO_2), and carbon dioxide (CO_2) can cause health problems, especially respiratory and cardiovascular diseases (Li et al., 2017; Zheng et al., 2013). This condition becomes even more serious in intersections and highway intersections that have a high level of vehicle density (Yildiz & Sucuoglu, 2025).

The research location is at a highway intersection with very dense vehicle activity during the morning rush hour, namely from 06:30 to 07:30 a.m., with a density of approximately 200 motorized vehicles per minute. The high mobility of vehicles during this period has the potential to significantly increase the concentration of air pollutants. This condition makes the intersection area a strategic location to monitor and predict air quality in real-time. According to (Garcia et al 2025), Areas with high levels of traffic density are one of the main sources of increased concentrations of PM2.5 and harmful gases in urban areas.

Manual air quality monitoring using conventional monitoring stations has several limitations, such as high installation costs, limited area coverage, and lack of ability to provide predictive information quickly (Bandara et al., 2021). The development of Internet of Things (IoT) technology provides a new solution in environmental monitoring systems through the use of intelligent sensors that are able to collect data continuously and in real time (Kumar et al., 2025). IoT technology allows the integration of various air quality sensors with the internet network so that data can be sent, stored, and analyzed automatically (Ramadan et al., 2024). The use of low-cost IoT-based sensors also allows the implementation of monitoring systems over a wider area at a more cost-efficient cost than traditional monitoring stations (Bandara et al., 2021). Although air quality monitoring is very important, the monitoring system alone is not enough to provide preventive measures against the impact of air pollution. A prediction system is needed that is able to estimate air quality conditions in the next period so that the community and the government can take mitigation steps early (Zheng et al., 2013). Air quality prediction is included in the category of time-series forecasting because air pollution data has a temporal

*) Coresponding Author: ahmadmujianto@unhasy.ac.id

pattern that is influenced by time, vehicle activity, temperature, humidity, and other environmental conditions (Talamanova & Pllana, 2022). Conventional statistical models such as ARIMA often have difficulty in capturing non-linear and complex relationships from air quality data (Talamanova & Pllana, 2022).

The development of Artificial Intelligence (AI) and Deep Learning methods has become a very interesting approach in air quality prediction research. One of the widely used methods is Long Short-Term Memory (LSTM), which is the development of the Recurrent Neural Network (RNN) designed to solve the problem of long-term dependency on sequential data (Zhang & Ding, 2017). LSTMs have the ability to store historical information over the long term so they are very suitable for predicting changes in air quality based on previous sensor data (Li et al., 2017). Research by (Zhang dan Ding, 2017) showed that the LSTM model is able to produce better accuracy in air quality predictions than traditional methods.

Several previous studies have also shown the effectiveness of deep learning methods in forecasting air quality. (Wang et al., 2018) develop a CNN-LSTM hybrid model to improve the accuracy of sensor data-driven air quality predictions. (Qi et al., 2019) stated that the combination of CNN and LSTM was able to improve the predictive performance of the Air Quality Index (AQI) on multivariate data. In addition (Zheng dan Zhang, 2019) develops a deep learning-based approach using SE-CNN to improve air pollution prediction performance. Others research by (Guo et al., 2025) also showed that the CNN-LSTM model performed well in predicting the PM2.5 index.

In addition to CNN-LSTM, the pure LSTM-based approach also showed excellent performance in air quality time-series data. (Keerthana et al., 2024) shows that the RNN and LSTM methods are able to provide stable prediction results on real-time air quality data. Sachetti and Mota (2021) through a systematic mapping study concluded that the LSTM-based model is one of the most widely used methods and has the best performance in time-series-based air quality prediction research.

Recent research also shows that the integration of IoT and deep learning can increase the effectiveness of air quality monitoring systems in real time. (Ramadan et al., 2024) develop an IoT-based AI system for automatic monitoring and forecasting of air pollution. Meanwhile, (Yildiz dan Sucuoglu, 2025) develop a real-time IoT-based air quality forecasting system using machine learning and obtain high predictive performance. Research (Garcia et al., 2025) also emphasized that the integration of IoT and AI is a major trend in the development of smart environmental monitoring systems.

In addition to conventional deep learning-based approaches, some recent research has also begun to develop more complex models. (Hettige et al., 2024) introduces AirPhyNet based on physics-guided neural network to improve the accuracy of air quality prediction. Meanwhile, (Faldo, 2024) developed a Jakarta air quality prediction system using LSTM and Bayesian Optimization which showed a significant improvement in model performance.

Based on various previous studies, the LSTM method was chosen in this study because it has an excellent ability to study temporal patterns of air quality data in real-time. This study proposes an IoT-based air quality prediction system using the Long Short-Term Memory (LSTM) method in highway intersections with high vehicle density. Air quality data is obtained from IoT sensors installed at the research site and used as an input LSTM model to predict air quality conditions over a given period of time. The proposed system is expected to be able to provide air quality prediction information quickly, accurately, and in real-time so that it can help the community and the government in mitigating the impact of air pollution in urban areas.

2. Theoretical Framework

2.1. Air Quality

Air quality is the air condition at a certain location that is measured based on the concentration of pollutants in the atmosphere. Air quality parameters generally include particulate matter (PM2.5 and PM10), carbon monoxide (CO), carbon dioxide (CO_2), nitrogen dioxide (NO_2), temperature, and air humidity. High concentrations of air pollutants can have a negative impact on human health, especially respiratory and cardiovascular system disorders (Li et al., 2017; Garcia et al., 2025).

In urban areas, the main source of air pollution comes from motor vehicle activities. Vehicle exhaust emissions produce pollutants that can increase the air pollution index, especially in areas with high traffic density such as highway intersections (Zheng et al., 2013). Therefore, air quality monitoring and prediction are very important to support early mitigation systems against air pollution (Yildiz & Sucuoglu, 2025).

2.2. Time Series Forecasting

Time-series forecasting is a method of prediction based on historical data patterns arranged sequentially over time (Talamanova & Pllana, 2022). In air quality research, IoT sensor data is categorized as time-series because pollutant values change periodically based on time and environmental conditions. The characteristics of time-series data of air quality area. have a temporal pattern, b. are influenced by time, c. are non-linear, and d. have dynamic fluctuations. Due to the complex nature of the data, deep learning methods were chosen to improve the accuracy of air quality predictions (Sachetti & Mota, 2021).

2.3. Deep Learning

Deep learning is a branch of machine learning that uses a stratified neural network to learn complex patterns from data (LeCun et al., 2015). Deep learning is highly effective in processing big data and air-quality time-series data (Li et al., 2017).

2.4. Long Short-Term Memory (LSTM)

Long Short-Term Memory is a development of the Recurrent Neural Network (RNN) designed to solve the problem of long-term dependency on sequential data (Hochreiter & Schmidhuber, 1997). LSTM is able to store historical information through a memory mechanism (cell state) so it is very suitable for use in time series based air quality prediction (Zhang & Ding, 2017).

2.5. Forget Gate

Forget gates are used to determine which information to retain or delete from previous memory (Hochreiter & Schmidhuber, 1997).

$$ft = \sigma(Wf[ht-1, xt] + bf) \quad (1)$$

- $ft = forget\ gate$
- $Wf = weight\ forget\ gate$
- $ht - 1 = Previous\ hidden\ state$
- $xt = Curent\ input$
- $bf = bias$
- $\sigma = sigmoid\ activation$

2.6. Input Gate

The input gate determines the new information to be stored in the cell state.

$$it = \sigma(Wi[ht-1, xt] + bi) \quad (2)$$

- $it = input\ gate$
- $Wi = bobot\ input\ gate$
- $bi = bias\ input\ gate$

2.7. Cell State

Cell state functions to store long-term information.

$$ct = ft \odot ct-1 + it \odot c{\sim}t \quad (3)$$

- $ct = New\ cell\ state$
- $ct - 1 = Previous\ cell\ state$
- $\odot = Element - wise\ multiplication$

2.8. Output Gate

The output gate generates the prediction output.

$$ht = ot \odot \tan h(ct) \quad (4)$$

- $ot = output\ gate$
- $Wo = bobot\ output\ gate$
- $bo = bias\ output\ gate$

The LSTM model is widely used in air quality forecasting because it has a high ability to study the temporal patterns of sensor data (Keerthana et al., 2024).

3. Method

3.1. Types of Research

This study uses an experimental research method with a quantitative approach. The research is focused on the development of a real-time air quality prediction system based on the Internet of Things and the Long Short-Term Memory (LSTM) method. A quantitative approach is used to analyze the data from the measurement of air quality sensors and evaluate the performance of the predictive model built.

3.2. Research Location and Time

The research was conducted in the area of highway intersections, markets, and sugar factories with high traffic density. Data collection was carried out during morning rush hour, namely 06.30 - 07.30 am in Cukir village, Diwek Jombang District, East Java, with an average vehicle density of around 200 motorized vehicles per minute. The research location was chosen because it has a high potential for air pollution due to motor vehicle activities.

3.3. Data Acquisition System

The data acquisition system uses an air quality sensor connected to an ESP32 microcontroller. The sensor is used to read air quality parameters in real-time. The measured parameters include:

1. Air Quality Index (AQI)
2. PM2.5
3. PM10
4. Air temperature
5. Air Humidity
6. CO_2 gas concentration

Sensor data is read periodically at a given interval and then sent to the server using the internet network.

Table 1 Data retrieved from the sensor

NO	CO	CO2	TEMPE-RATURE	HUMIDITY	AQI
1	3.113	600	31.6	57.5	38.9
2	3.104	586	31.7	57.5	38.8
3	3.037	599	31.7	57.9	38
4	3.104	581	31.7	57.6	38.8
5	3.050	578	31.6	58.1	38.1
6	3.117	603	31.6	57.9	39
7	3.105	597	31.5	57.8	38.8
8	3.116	589	31.6	58	39
9	3.062	583	31.4	58.2	38.3
10	3.120	581	31.5	58.5	39
11	3.076	578	31.5	58.4	38.5
12	3.112	584	31.4	58.7	38.9
13	3.056	601	31.3	58.6	38.2
14	3.075	588	31.3	58.9	38.4
..	…	….	….	….	….

NO	CO	CO2	TEMPE-RATURE	HUMIDITY	AQI
235	0.318	419	23.6	85.7	4
236	0.321	404	23.5	86.1	4
237	0.317	416	23.5	86.2	4
238	0.320	430	23.5	86.1	4
239	0.310	424	23.5	86.5	3.9

4. Results and Discussion

4.1. Results

The prediction results show that the LSTM model is able to follow the pattern of changes in air quality based on the sensor's historical data. The MAE and RMSE values showed a relatively small rate of prediction error, while the R^2 values showed a fairly good model in studying the temporal patterns of AQI. The LSTM model is effectively used for real-time air quality forecasting in high-density vehicle conditions.

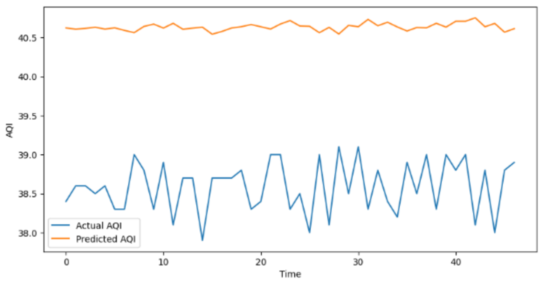

Figure 1. AQI Prediction Results

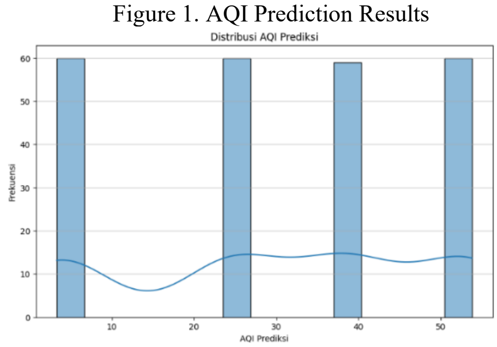

Figure 2 Predicted AQI Distribution

Proportion of air quality

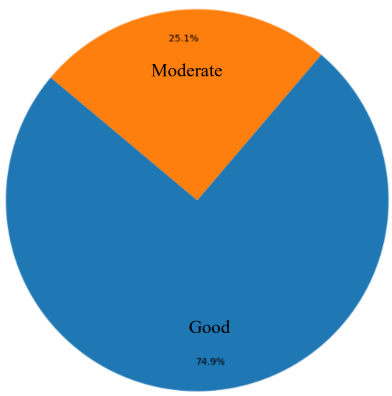

Figure 3 Proportion of Air Quality Categories

'AQI_Prediksi' Histogram Analysis: The histogram shows the distribution of 'AQI_Prediksi' values. From the plot, we can see where most of the values are **concentrated**, the shape of their distribution (for example, whether it's sloping or symmetrical), and the range of minimum to maximum values. This provides an overview of the dominant air quality characteristics in your dataset. **'Air Quality Category' Pie Chart Analysis:** This pie chart depicts the relative proportions of each air quality category (e.g. 'Good', 'Moderate', 'Unhealthy'). These visualizations help you understand which categories appear most frequently, show the most common air quality conditions, and assess whether the data is evenly distributed among categories or if there are significant class imbalances. For your data, it can be seen that the 'Good' category is very dominant compared to 'Medium' (179 vs 60). This visualization complements previous MAPE and MSE metrics, providing a more comprehensive understanding of the model's data and performance

Based on the results of your LSTM model's evaluation for the 'AQI_Prediksi' column: **Mean Squared Error (MSE): 3.5407** This MSE value represents the mean square of the difference between the actual 'AQI_Prediksi' value and the value predicted by the model. The smaller the MSE value, the better the model match with the data. The value of 3.5407 indicates that there are some errors in the prediction, however whether or not this is acceptable depends largely on the scale of the original data. **Mean Absolute Percentage Error (MAPE): 48.37%.** A MAPE of 48.37% means, on average, your model's prediction deviates about 48.37% from the actual value. Based on the automated analysis I provided, it is categorized as a 'moderate MAPE (20-50%)', which indicates the model has acceptable prediction accuracy but there may be room for improvement.

4.1. Discussion

This study aims to build a real-time air quality prediction system using Internet of Things based sensors and the Long Short-Term Memory (LSTM) method in highway intersections with high vehicle density. The research location has a vehicle intensity of around 200 motorized vehicles per minute during morning rush hour, namely 06.30–07.30 WIB. High vehicle activity causes an increase in exhaust emissions which has an impact on dynamic air quality fluctuations.

Based on the results of the model test, the LSTM method was able to predict the value of the Air Quality Index (AQI) quite well against the pattern of changes in air quality sensor data. This can be seen from the prediction results that are able to follow the actual data trends at most **observation** time intervals. The model's ability to study temporal patterns shows that the LSTM method is effectively used on dynamic and continuous air quality time-series data.

The Mean Absolute Error (MAE) value of 2,046 indicates that the average difference between the prediction results and the actual data is relatively small. A Root Mean Square Error (RMSE) value of 2,076 also indicates that the model's prediction error rate is still within the acceptable range for a real-time air quality monitoring system. The low error value indicates that the model is able to study the relationships between historical air quality data quite well.

However, the test results show that the R^2 Score value is still **negative**. This condition indicates that the model is not fully optimal in explaining the overall variation in air quality data. This may be due to several factors, including:

1. the number of datasets that is still limited,
2. relatively short duration of data collection,
3. variations in environmental parameters that are not too complex,
4. as well as very fast fluctuations in air quality during peak vehicle hours.

Time-series data on air quality has non-linear characteristics and is greatly influenced by environmental conditions such as temperature, humidity, wind direction, and vehicle volume. Therefore, the LSTM model requires a larger amount of historical data in order to be able to study temporal patterns more optimally (Sachetti & Mota, 2021).

However, the results of this study show that the integration of IoT and LSTM can be used as a real-time air quality monitoring and forecasting solution in urban environments. The developed system is capable of automatically capturing sensor data and generating air quality predictions without manual intervention. This approach supports the development of a smart environmental monitoring system that is more efficient and adaptive than conventional monitoring systems.

The results of this study are in line with the research conducted by (Zhang and Ding, 2017) which stated that the LSTM method has a high ability to predict air quality based on time series data. Research (Li et al., 2017) It also shows that deep learning performs better than traditional statistical methods in the case of air quality prediction. In addition, the research (Talamanova dan Pllana, 2022) stated that LSTM is more effective than the ARIMA method in studying temporal patterns of real-time air quality data.

Compared to previous research, this study has several aspects of novelty. First, this study uses real-time data obtained directly from IoT sensors at highway intersections with high vehicle density. Most previous studies have used public datasets or government monitoring station data (Zheng et al., 2013; Li et al., 2017). The use of sensor data directly contributes to the development of an actual low-cost IoT-based air quality monitoring system in the field. Second, this study focuses on the observation of air quality during morning rush hour with the characteristics of heavy traffic and very rapid changes in pollutants. This approach contributes to the analysis of air quality based on micro traffic conditions, which have been relatively little discussed in previous studies.

Third, this study integrates monitoring and forecasting systems in one real-time platform based on IoT and deep learning. Previous research has generally only focused on monitoring or predicting separately (Bandara et al., 2021). The integration of the two systems allows the development of an early warning system against air pollution in urban areas.

In addition, the use of the LSTM method in this study shows the ability of deep learning in studying air quality patterns based on low-cost sensor data. This supports research (Garcia et al., 2025) which states that the combination of IoT and Artificial Intelligence is the main trend in the development of smart cities and environmental monitoring systems.

Overall, this study shows that the LSTM method has good potential for use in IoT-based real-time air quality prediction systems. For further research, model performance can be improved by:

1. increasing the number of datasets,
2. prolong the duration of observation,
3. add meteorological parameters,
4. using hybrid models such as CNN-LSTM or BiLSTM,
5. as well as optimizing the hyperparameter model.

Thus, the developed system has the potential to be a supporting solution in urban air quality monitoring and early mitigation of the impact of air pollution due to motor vehicle activities.

5. Conclusion

This research has succeeded in developing a real-time air quality prediction system based on the Internet of Things and the Long Short-Term Memory (LSTM) method in highway intersections with high vehicle density. The results of the study show that the LSTM method is able to study the temporal patterns of air quality data and produce Air Quality Index (AQI) predictions quite well. Based on the results of the evaluation, the model obtained a Mean Absolute Error (MAE) of 2,046 and a Root Mean Square Error (RMSE) of 2,076, indicating a relatively small level of prediction error compared to the actual data. The integration of IoT and LSTM technology has proven to be effective in supporting automatic and real-time air quality monitoring and forecasting systems. The novelty aspect of this research lies in the use of real-time sensor data in heavy traffic areas as well as the integration of air quality monitoring and prediction in a deep learning-based system. The developed system has the potential to be used as a smart environmental

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/84892

monitoring and early warning system to help mitigate air pollution in urban areas.

Bibliography

Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T., 2017. Deep learning architecture for air quality predictions. Environmental Science and Pollution Research, 24(22), 22453–22464.

Zhang, Y., & Ding, W., 2017. Air quality prediction based on LSTM-RNN model. IEEE 2nd International Conference on Big Data Analysis (ICBDA), 506–511.

Wang, Y., Li, J., & Li, L., 2018. A hybrid model for air quality prediction based on CNN and LSTM. IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 119–124.

Qi, R., Li, Q., & Li, Y., 2019. A hybrid model for air quality index prediction using CNN and LSTM. International Journal of Environmental Research and Public Health, 16(17), 3167.

Zheng, Y., Liu, F., & Hsieh, H.P., 2013. U-Air: When urban air quality inference meets big data. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1436–1444.

Zheng, H., & Zhang, Y., 2019. A deep learning approach for air quality prediction based on SE-CNN. IEEE Access, 7, 89474–89483.

Talamanova, I., & Pllana, S., 2022. Data-driven real-time short-term prediction of air quality: Comparison of ES, ARIMA, and LSTM. arXiv preprint arXiv:2211.09814.

Sachetti, L. L. S., & Mota, V. F. S., 2021. Time series prediction about air quality using LSTM-based models: A systematic mapping. arXiv preprint arXiv:2111.11848.

Ramadan, M. N. A., 2024. Real-time IoT-powered AI system for monitoring and forecasting air pollution. Computerized Chemical Engineering, Elsevier.

Kumar, P. R., 2025. Air quality prediction using IoT and neural networks: An empirical analysis. Journal of Theoretical and Applied Information Technology, 103(3), 1076–1082.

Yildiz, O., & Sucuoglu, H. S., 2025. Development of real-time IoT-based air quality forecasting system using machine learning approach. Sustainability, 17(19), 8531.

Garcia, A., 2025. Advancements in air quality monitoring: A systematic review of IoT-based systems integrated with artificial intelligence. Artificial Intelligence Review.

Guo, Z., Wu, S., Zhu, M., & Guandi, H., 2025. Air Quality PM2.5 Index Prediction Model Based on CNN-LSTM. arXiv preprint arXiv:2508.11215.

Faldo, R., 2024. APD-BayTM: Jakarta Air Quality Index Prediction using LSTM and Bayesian Optimization. Jurnal Nasional Teknik Elektro dan Teknologi Informasi.

Keerthana, G., 2024. Air Quality Prediction Using RNN and LSTM. Journal of Information Technology.

Bandara, N., Hettiarachchi, S., & Athukorala, P., 2021. AirSPEC: An IoT-empowered air quality monitoring system integrated with machine learning framework. arXiv preprint arXiv:2111.14125.

Hettige, K. H., 2024. AirPhyNet: Harnessing physics-guided neural networks for air quality prediction. arXiv preprint arXiv:2402.03784.