

Application of Long Short Term Memory, Gated Recurrent Unit, and Hybrid LSTM-GRU in Air Quality Forecasting Based on Air Pollution Index in Semarang City

Nurul Hidayah*, Putriaji Hendikawati

Department of Mathematics, Faculty of Mathematics and Natural Science,
Universitas Negeri Semarang, Semarang, 50229, Indonesia

Article Info

Article History:

Received 04-07-2025

Accepted 20-10-2025

Published 31-10-2025

Keywords:

*LSTM, GRU, Hybrid LSTM
GRU, Forecasting, Air
Pollution Index*

Abstract

Air pollution harms human health, and Semarang City is among the most polluted cities in Indonesia, so forecasts of the Air Pollution Index (ISPU) are needed to anticipate poor air quality. This study compares three deep learning methods for forecasting the ISPU: Long Short Term Memory (LSTM), which retains long-term information through its memory cell and gates and overcomes the vanishing gradient problem; Gated Recurrent Unit (GRU), which has fewer gates and parameters and therefore trains faster; and a Hybrid LSTM-GRU, which is expected to combine the long-term memory of LSTM with the efficiency of GRU. The data are 35,064 hourly observations of seven pollutant parameters (PM10, PM2.5, SO2, CO, O3, NO2, and HC) from the Semarang City Environmental Agency, from January 1, 2021 to December 31, 2024. After time-based interpolation of missing values, min-max normalization, an 80:20 split, and a sliding window of 24 hours, each model was trained with two numbers of neurons and three learning rates (six combinations per model) and evaluated with the mean RMSE and MAE. The best LSTM (100 neurons, learning rate 0.001) gave a mean RMSE of 0.7631 and MAE of 0.2470, the best GRU (50 neurons, 0.001) 0.7600 and 0.2215, and the best Hybrid LSTM-GRU (50 neurons, 0.001) 0.7597 and 0.2193. The Hybrid model is the most accurate, but its advantage over GRU is very small (0.04% in RMSE and 1.0% in MAE) and is clearer against LSTM in MAE (11.2%). The 24-hour forecast for January 1, 2025 shows PM2.5 as the determining parameter, with unhealthy air quality in the early morning and late evening. The forecasts can support early warning; the models should still be tested with repeated runs and against baseline models.

How to cite:

Hidayah, N., & Hendikawati, P. (2025). Application of long short term memory, gated recurrent unit, and hybrid LSTM-GRU in air quality forecasting based on air pollution index in Semarang City. *UNNES Journal of Mathematics*. 14(2): 29-38

1. Introduction

Forecasting is the process of predicting future events based on data or information from the past. The goal of forecasting is to generate predictions that are as accurate as possible and to provide supporting information in decision making. Forecasting can be applied in various fields, such as business, industry, finance, economics, politics, environmental science, medicine, social sciences, and government. Along with technological advancements, artificial intelligence currently plays a very important role in the forecasting process. AI has emerged as an important tool to offer effective and efficient solutions to various problems faced by humans (Farwati et al., 2023). Machine learning is a part of artificial intelligence that uses statistical methods so that computers can learn and improve their capabilities automatically (Subramaniam et al., 2022). Deep learning is a part of machine learning that plays an important role in fields that use large amounts of high-dimensional data. Deep learning is capable of surpassing simpler machine learning algorithms in processing various types of data, such as text, audio, video, images, and more. Additionally, deep learning can also be applied in the processing of sequential data including time series data (Saputra, 2023). Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are deep learning methods that can be used to analyze and predict time series data.

Air is a natural resource that plays an important role in the survival of all living beings on earth, including humans, animals, and plants. Therefore, good air quality needs to be maintained to keep the air clean and free from harmful substances that can endanger health. However, air pollution can lead to a decrease in air quality. Air pollution occurs due to the entry of harmful substances or components into the atmosphere that cause the quality of air to decline. This pollution can be caused by natural factors, such as volcanic eruptions and forest fires. In addition, human activities such as urbanization and industrial growth also contribute significantly to air pollution. The impact of air pollution is very dangerous to health, as it can cause various diseases, such as respiratory disorders, lung cancer, stroke, heart disease, and even death. The State of Global Air 2024 report states that air pollution caused 8.1 million deaths worldwide in 2021, with approximately 58% of those deaths attributed to $\text{PM}_{2.5}$ pollution. This figure makes air pollution the second largest risk factor for death globally, including among children under five years old. Of the total deaths, non-communicable diseases such as heart disease, stroke, diabetes, lung cancer, and chronic obstructive pulmonary disease (COPD) contributed to almost 90% of the disease burden caused by air pollution. In the same year, over 700,000 deaths in children under 5 years old were attributed to air pollution. This number accounts for about 15% of the total global deaths in that age group (Health Effects Institute, 2024). The problem of air pollution is also faced in various major cities in Indonesia, including Semarang City. The IQAIR report states that in 2024, Semarang City ranked 17th as the city with the highest pollution levels in Indonesia (IQAir, 2025). One of the causes of air pollution in Semarang City is the increasing number of motor vehicles. According to data from the BPS-Statistics Jawa Tengah Province in 2024, the number of motor vehicles in Semarang City reached 1,985,109 units. This number represents an increase of 67,831 units from 2023. Emissions from motor vehicles produce air pollutants, such as carbon dioxide (CO_2), carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen oxides (NO_x), which can be harmful to human health (Rahmat et al., 2023).

Air quality refers to the state of the air that is inhaled by humans and other living beings. Air quality can be analyzed through the Air Pollution Index, which is a unitless number that describes the ambient air quality conditions in a specific area. This number is based on its impact on human health, aesthetic value, and other living beings. The ISPU serves as a parameter to measure air quality by indicating the level of air pollution originating from chemical substances and particles contained within it (Perdana & Muklason, 2023). There are seven parameters used in the ISPU calculation, namely particulate matter 10 (PM_{10}), particulate matter 2.5 ($\text{PM}_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and hydrocarbons (HC). Monitoring air quality is one way to determine the level of air pollution. Therefore, to understand the future condition of air quality, a forecasting method is required. Air quality forecasting plays an important role in efforts to control air pollution. Accurate forecasting outcomes can provide an overview of future air pollution conditions, allowing for anticipatory measures against the impacts caused by air pollution (Koesoemaningroem et al., 2021).

The urgency of this research follows from the health burden of air pollution and from the lack of short-term forecasts for Semarang City. High and rising pollutant levels are a threat now, and the ISPU is reported as an index for the public; a forecast for the next hours allows residents, especially vulnerable groups, and the authorities to take precautions before air quality deteriorates, whereas monitoring alone shows only what has already happened. Accurate hourly forecasts of all seven ISPU parameters are therefore needed.

LSTM is one of the deep learning methods developed from Recurrent Neural Network (RNN). Hochreiter and Schmidhuber first introduced the LSTM architecture in 1997. LSTM is specifically designed to address the vanishing gradient problem in RNN models when learning long-term data. LSTM has three gates, namely the forget gate, input gate, and output gate. LSTM also has a memory cell that can store information for a long time, making it effective for processing time series data. With greater memory capacity, LSTM can process and retain previous information, allowing this method to perform better when dealing with large amounts of data. GRU is a deep learning algorithm derived from the development of LSTM. This method has a simpler structure consisting of two gates, namely the reset gate and the update gate. Therefore, its computation process is faster compared to LSTM due to having fewer parameters. By considering the advantages of each method, the combination of LSTM and GRU into a Hybrid model, namely Hybrid LSTM-GRU, aims to combine the ability of LSTM to store long-term information with the computational speed of GRU. This Hybrid approach is expected to produce more optimal and accurate predictions. The research conducted by Yurtsever (2023) on forecasting unemployment rates using the Hybrid LSTM-GRU method shows that the Hybrid method is superior to LSTM and GRU

methods separately. Deep learning approaches such as LSTM, GRU, and Hybrid LSTM-GRU can be applied in the field of environmental science, particularly in forecasting air quality.

LSTM, GRU, and the Hybrid LSTM-GRU were chosen for three reasons. First, hourly ISPU data are sequential, nonlinear, and depend on both recent and earlier hours, which recurrent networks with gates can learn, while the memory cell of LSTM (Hochreiter & Schmidhuber, 1997) overcomes the vanishing gradient problem of ordinary RNNs. Second, GRU (Cho et al., 2014) has fewer gates and parameters than LSTM, so it trains faster and may be as accurate on data of moderate complexity. Third, the Hybrid LSTM-GRU is expected to keep the long-term memory of LSTM and the efficiency of GRU in one model, and it has outperformed single models in other fields (Tripathy et al., 2023; Yurtsever, 2023). Because no method is known in advance to be best for ISPU data, the three are compared under the same data, preprocessing, hyperparameter grid, and evaluation metrics.

Previous studies of air quality forecasting in Indonesia have mostly used statistical or machine learning methods: ARIMA for air quality in Semarang (Khusna et al., 2023), DSARIMA for Surabaya (Koesoemaningroem et al., 2021), support vector regression for the air quality index in Makassar (Rahmat et al., 2023), and ARCH/GARCH and LSTM for PM_{2.5} in Central Jakarta, where LSTM was superior to GARCH (Fauzan et al., 2024). A hybrid ARIMAX-LSTM has been applied to the ISPU of DKI Jakarta (Perdana & Muklason, 2023). Comparisons among recurrent networks come from other fields: GRU was more accurate than RNN and LSTM for LQ45 stock prices (Nilsen, 2022), and the LSTM-GRU hybrid outperformed LSTM and GRU for unemployment rates (Yurtsever, 2023) and stock market indices (Tripathy et al., 2023). To the authors' knowledge, the Hybrid LSTM-GRU has not yet been compared with LSTM and GRU for the ISPU of Semarang City.

The novelty of this study is therefore threefold: it applies and compares LSTM, GRU, and Hybrid LSTM-GRU on four years of hourly data of all seven ISPU parameters of Semarang City; it evaluates every model on the same grid of neurons and learning rates, so that the effect of each hyperparameter on each architecture can be seen; and it translates the forecast of the best model into ISPU categories for the next 24 hours.

This study aims to compare the LSTM, GRU, and Hybrid LSTM-GRU methods in forecasting air quality based on the Air Pollution Index in Semarang City. The research questions are: (1) which combination of the number of neurons and the learning rate gives the best architecture for each method; (2) which of the three methods is the most accurate according to RMSE and MAE; and (3) what air quality category does the best model forecast for the next 24 hours?

2. Method

The data used in this research is the Air Pollution Index data in Semarang City obtained from the Semarang City Environmental Agency. The ISPU data consists of seven air pollutant parameters, namely particulate matter 10 (PM₁₀), particulate matter 2.5 (PM_{2.5}), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), ozone (O₃), and hydrocarbons (HC). This data was collected during the period from January 1, 2021 to December 31, 2024. The data analysis techniques in this research were carried out through several stages so that the data used is ready for analysis and can produce optimal predictions. The stages of data analysis in this research are as follows.

1. Data Collection
Collecting data on the Air Pollution Index (ISPU) from the Semarang City Environment Agency.
2. Data Preprocessing
Data preprocessing aims to prepare raw data for analysis. This research has four stages of preprocessing, as follows:
 - a. Data cleaning
At this stage, a check is performed for missing values to determine whether the research data has any empty or missing values. If the data has missing values, it needs to be handled by filling in those missing values using interpolation methods.
 - b. Data normalization
Data normalization aims to adjust the value of the data into the same range, which is from 0 to 1. The normalization method used in this research is Min-Max Scaling.
 - c. Data splitting
The research data is divided into two parts: 80% training data and 20% testing data.
 - d. Sliding window
The sliding window aims to transform the data into a sequential form.
3. Modelling
Create the architecture for LSTM, GRU, and Hybrid LSTM-GRU models to obtain the best model. The models are built by combining hyperparameters. The hyperparameters tested in this study include the number of neurons and learning rate. Each model will undergo several tests based on the predetermined number of neurons and learning rate to achieve the model with the best hyperparameters.
The Hybrid LSTM-GRU model combines both layer types in one network, so that the LSTM component retains long-term information and the GRU component adds computational efficiency. Two numbers of neurons (50 and 100) and three learning rates (0.1, 0.01, and 0.001) give six combinations per model and 18 trained models in total.
4. Model Testing
Conduct model testing for LSTM, GRU, and Hybrid LSTM-GRU based on the best model architecture. This model is used to predict the Air Pollution Index (ISPU) values on the testing data. At this stage, the data is still in its normalized

form, so a denormalization process is performed to return the data to its original scale for easier understanding and comparison with the actual data.

5. Model Evaluation

Conducting a model evaluation to measure the performance of the LSTM, GRU, and Hybrid LSTM-GRU models. From the predictions of this data, the accuracy level of each model can be determined. The performance measurement of the model is done using evaluation metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). After obtaining the model evaluation results, a comparison of the accuracy of each model is carried out to determine the best method in this research.

For n test observations with actual values y_i and predicted values $\hat{y}_i$, the metrics are

$$\text{RMSE} = \sqrt{[(1/n) \sum (y_i - \hat{y}_i)^2]}, \quad \text{MAE} = (1/n) \sum |y_i - \hat{y}_i|$$

RMSE penalizes large errors more than MAE, so the two metrics together show both the typical error and the effect of extreme values. The errors are expressed in ISPU units after denormalization, and the reported mean is taken over the seven pollutant parameters.

6. Forecasting

Forecasting air quality in the city of Semarang for the next 24 hours based on the best method.

The forecast is read with the ISPU categories of the Indonesian regulation: 1–50 good, 51–100 moderate, 101–200 unhealthy, 201–300 very unhealthy, and above 300 hazardous (Ministry of Environment and Forestry, 2020).

3. Results and Discussion

3.1. Data Collection

The data used in this study is the Air Pollution Index (ISPU) data of Semarang City, which consists of seven air pollutant parameters, namely particulate matter 10 (PM_{10}), particulate matter 2.5 ($\text{PM}_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and hydrocarbons (HC). The data were collected over a period from January 1, 2021 at 00:00 to December 31, 2024 at 23:00, with a total of 35,064 data point for each air pollutant parameter.

3.2. Data Preprocessing

There are four stages that are carried out, namely data cleaning, data normalization, data splitting, and sliding window.

3.2.1. Data Cleaning

At this stage, a process is conducted to check for missing data or missing values.

```
print("Jumlah Missing Value:")
print(data.isnull().sum())
```

Jumlah Missing Value:
 PM10 3888
 PM2.5 3905
 SO2 4025
 CO 4149
 O3 5075
 NO2 5235
 HC 13513
 dtype: int64

Figure 1. Check missing value

The results of the inspection show that there are missing values in every air pollutant parameter. Therefore, it is necessary to address the missing values by using a time-based interpolation method. This method uses the time information between rows to estimate values between known data points.

3.2.2. Data Normalization

The next stage is data normalization aimed at adjusting the scale of the data to the same range, which is from 0 to 1. The normalization technique used is the Min-Max Scaling method. The results of the normalization are presented in Table 1.

Table 1. Data Normalization

Time	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
2021-01-01 00:00:00	0,1250	0,1158	0,9032	0,0923	0,16	0,2625	0
2021-01-01 01:00:00	0,1364	0,1237	0,9194	0,0923	0,16	0,2625	0
2021-01-01 02:00:00	0,1477	0,1316	0,9355	0,0923	0,16	0,2625	0
...	...	...	...	...	...	...	...
2024-12-31 21:00:00	0,1818	0,5684	0,3871	0,0308	0,10	0,9625	0
2024-12-31 22:00:00	0,1705	0,5553	0,3710	0,0308	0,10	0,9625	0
2024-12-31 23:00:00	0,1591	0,4921	0,3710	0,0308	0,10	0,9625	0

3.2.3. Data Splitting

The next step is to divide the data into two parts, namely training data and testing data. This division is performed to train the model using the training data, then test the model's performance using the testing data that has never been seen before. This study uses 80% training data with a total of 28,051 data points and 20% testing data with a total of 7,013 data points for each variable.

3.2.4. Sliding Window

Sliding window is a technique used to transform time series data into a sequential form. This technique organizes the sequential data into input-output pairs, with a number of previous data points equal to the time step as input (X) and the subsequent data point as output (Y). The length of the time step used in this study is 24, which means that the ISPU data for the last 24 hours is used as input to predict the ISPU value at the next time step, which is the 25th data.

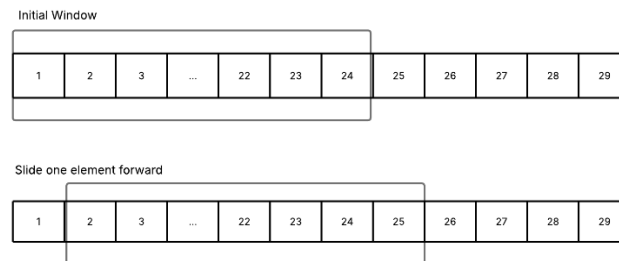


Figure 2. Sliding Window

3.3. Modelling

At this stage, LSTM, GRU, and Hybrid LSTM-GRU models will be built by combining hyperparameters to get the best model architecture. The parameters to be tested in this research are the number of neurons and learning rate.

Table 2. Hyperparameter Combination

Parameter	Value
Neuron	50, 100
Learning Rate	0,1, 0,01, 0,001

After testing each model, the results are shown in Table 3.

Table 3. Comparison of Hyperparameter Testing

Neuron	Learning Rate	LSTM		GRU		Hybrid LSTM-GRU	
		Mean of RMSE	Mean of MAE	Mean of RMSE	Mean of MAE	Mean of RMSE	Mean of MAE
50	0,001	0,7762	0,2511	0,7600	0,2215	0,7597	0,2193
50	0,01	0,7880	0,2843	0,7946	0,2787	0,7900	0,2573
50	0,1	13,5535	9,1375	10,8002	3,3401	6,9353	3,2159
100	0,001	0,7631	0,2470	0,7621	0,2303	0,7609	0,2337
100	0,01	0,7721	0,2601	0,8006	0,2923	0,8157	0,3865
100	0,1	11,3150	6,8750	15,6051	6,6277	9,4770	5,5895

Based on the hyperparameter combination results, the learning rate value of 0.001 produces the lowest error values in all models. When the learning rate value is increased to 0.01, the model performance tends to decrease as indicated by an increase in the mean RMSE and MAE values. At a learning rate of 0.1, all models experienced a very significant increase in error values. Meanwhile, increasing the number of neurons from 50 to 100 does not always improve model accuracy. In some tests, it actually causes an increase in the error value. Based on the test results, the best architecture is obtained for the LSTM model with 100 neurons and a learning rate of 0.001, the GRU model with 50 neurons and a learning rate of 0.001, and the Hybrid LSTM-GRU model with 50 neurons and a learning rate of 0.001. These results are consistent with research by Yulisa et al. (2023) on forecasting the value of oil and gas exports in Indonesia, which shows that the use of a learning rate of 0.001 is more optimal than learning rates of 0.01 and 0.1 for both LSTM and GRU models. In the study, the use of a larger number of neurons resulted in the best performance in the LSTM model. In contrast, the best GRU model was obtained with a smaller number of neurons. The best LSTM model used 20 neurons, while the best GRU model used 10 neurons. Therefore, it can be concluded that increasing the number of neurons does not consistently improve the accuracy of the model. However, the use of a smaller learning rate generally provides better performance. This is shown by the lower mean RMSE and MAE values.

The error ratios give a clearer picture of these effects. At 50 neurons, raising the learning rate from 0.001 to 0.1 multiplies the mean RMSE by about 17.5 for LSTM (0.7762 to 13.5535), 14.2 for GRU (0.7600 to 10.8002), and 9.1 for the Hybrid model (0.7597 to 6.9353); at 100 neurons the Hybrid model again has the lowest error at a learning rate of 0.1 (RMSE 9.4770, against 11.3150 for LSTM and 15.6051 for GRU). The Hybrid model is thus the least sensitive to an excessively large learning rate. The effect of the number of neurons differs by architecture: at learning rates of 0.001 and 0.01, LSTM is more accurate with 100 neurons (RMSE 0.7631 against 0.7762, and 0.7721 against 0.7880), whereas GRU and the Hybrid model are more accurate with 50 neurons (GRU 0.7600 against 0.7621 and 0.7946 against 0.8006; Hybrid 0.7597 against 0.7609 and 0.7900 against 0.8157). A plausible reason is that a GRU unit has fewer weights than an LSTM unit, so GRU-based layers need fewer neurons to represent the same patterns, while a larger number of neurons raises the risk of overfitting; this was not tested directly. The same pattern, a larger number of neurons for LSTM and a smaller one for GRU, was reported by Yulisa et al. (2023).

3.4. Model Testing

The best model architecture for each model that has been obtained in model training. Furthermore, it is used to test the model. This test is carried out using testing data. The results of model testing are shown in Table 4 to Table 6.

Table 4. LSTM Model Testing Results

Time	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
2024-03-15 19:00:00	2,191	-0,154	28,040	2,021	56,088	34,840	-0,025
2024-03-15 20:00:00	2,006	-0,154	28,041	2,021	51,952	34,847	-0,025
2024-03-15 21:00:00	1,992	-0,154	28,041	2,021	46,550	34,853	-0,025
...	...	...	...	...	...	...	...
2024-12-31 21:00:00	16,603	220,746	24,046	2,941	10,018	76,624	-0,025
2024-12-31 22:00:00	15,561	205,964	24,047	1,812	10,018	76,624	-0,025
2024-12-31 23:00:00	14,584	204,523	22,922	1,755	10,018	76,624	-0,025

Table 5. GRU Model Testing Results

Time	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
2024-03-15 19:00:00	2,057	-0,044	28,032	1,991	56,283	35,088	0,008
2024-03-15 20:00:00	1,978	-0,044	28,031	1,991	52,250	35,088	0,008
2024-03-15 21:00:00	1,961	-0,044	28,030	1,991	47,172	35,088	0,008
...	...	...	...	...	...	...	...
2024-12-31 21:00:00	16,591	218,653	24,030	2,959	9,340	76,828	0,008
2024-12-31 22:00:00	15,593	204,008	24,034	1,827	9,340	76,828	0,008
2024-12-31 23:00:00	14,600	202,240	23,026	1,781	9,340	76,828	0,008

Table 4. Hybrid LSTM-GRU Model Testing Results

Time	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
2024-03-15 19:00:00	1,979	-0,047	27,988	2,067	55,967	35,037	0,004
2024-03-15 20:00:00	1,946	-0,047	27,987	2,067	51,729	35,044	0,004
2024-03-15 21:00:00	1,943	-0,047	27,986	2,067	46,308	35,049	0,004
...	...	...	...	...	...	...	...
2024-12-31 21:00:00	16,443	219,572	23,965	3,046	9,263	76,609	0,004
2024-12-31 22:00:00	15,442	204,441	23,972	1,942	9,263	76,609	0,004
2024-12-31 23:00:00	14,499	203,197	22,796	1,830	9,263	76,609	0,004

3.5. Model Evaluation

The next stage is model performance evaluation to compare model accuracy values and get the best model in the study. Model performance assessment is carried out using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) evaluation metrics. The smaller the RMSE and MAE values of a model, indicating that the model is more optimal. The model with the smallest RMSE and MAE values is selected as the best model.

Table 7. Model Evaluation

	LSTM	GRU	Hybrid LSTM-GRU
Mean of RMSE	0,7631	0,7600	0,7597
Mean of MAE	0,2470	0,2215	0,2193

The model evaluation results show that the LSTM model obtained a mean RMSE value of 0.7631 and MAE of 0.2470. The GRU model obtained a mean RMSE value of 0.7600 and MAE of 0.2215. Meanwhile, the Hybrid LSTM-GRU model obtained a mean RMSE value of 0.7597 and MAE of 0.2193. Based on these results, it can be concluded that the most optimal model in this study is the Hybrid LSTM-GRU model. The results of this study are in accordance with the research of Tripathy et al. (2023) regarding stock market index forecasting which shows that the LSTM-GRU method is superior to the LSTM and GRU methods.

The advantage of the Hybrid model is small. Relative to GRU, its mean RMSE is lower by 0.04% (0.7597 against 0.7600) and its MAE by 1.0% (0.2193 against 0.2215); relative to LSTM, the RMSE is lower by 0.45% and the MAE by 11.2%. The three models are therefore practically equal in RMSE, and the Hybrid and GRU models differ from LSTM mainly in MAE. The RMSE is about three times the MAE for all models (3.09 for LSTM, 3.43 for GRU, and 3.46 for Hybrid), which means that the error is dominated by a few large deviations, probably sudden pollution spikes, and that all three models handle these extreme values similarly while GRU and Hybrid reduce the typical error more than LSTM.

3.6. Forecasting

After obtaining the most optimal model in the study, namely the Hybrid LSTM-GRU model. Next, air quality forecasting is carried out for the next 24 hours using the model. The forecasting process produces predictions gradually in the next hour until it reaches 24 hours. The prediction time range starts one hour after the time of the last data, which is January 1, 2025 at 00:00 to 23:00. The results of forecasting the next 24 hours for each air pollutant parameter are shown in Table 8.

Table 8. Forecasting Results for the Next 24 Hours

January 1, 2025	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
00:00:00	13,59	172,27	22,90	1,87	9,96	76,61	0,00
01:00:00	13,30	160,51	22,77	1,82	9,92	76,21	0,01
02:00:00	13,06	149,93	22,66	1,81	9,89	75,87	0,01
03:00:00	12,86	140,38	22,58	1,82	9,85	75,52	0,01
04:00:00	12,64	131,88	22,50	1,86	9,80	75,17	0,02
05:00:00	12,41	124,41	22,44	1,91	9,75	74,80	0,02
06:00:00	12,18	117,93	22,37	1,97	9,70	74,44	0,03
07:00:00	11,98	112,38	22,32	2,04	9,65	74,07	0,03
08:00:00	11,82	107,71	22,26	2,12	9,59	73,70	0,04
09:00:00	11,71	103,87	22,21	2,19	9,54	73,33	0,04
10:00:00	11,66	100,81	22,16	2,28	9,48	72,97	0,04
11:00:00	11,66	98,45	22,11	2,36	9,42	72,60	0,05
12:00:00	11,69	96,74	22,07	2,45	9,36	72,25	0,05
13:00:00	11,75	95,60	22,02	2,53	9,30	71,89	0,06
14:00:00	11,82	94,95	21,98	2,62	9,24	71,55	0,06
15:00:00	11,89	94,75	21,93	2,71	9,18	71,21	0,07
16:00:00	11,95	94,91	21,89	2,80	9,11	70,88	0,07
17:00:00	12,01	95,37	21,85	2,89	9,05	70,55	0,07
18:00:00	12,06	96,08	21,81	2,99	8,99	70,24	0,08
19:00:00	12,10	96,98	21,76	3,08	8,93	69,93	0,08
20:00:00	12,13	98,04	21,72	3,17	8,86	69,62	0,09
21:00:00	12,17	99,23	21,68	3,26	8,80	69,33	0,09
22:00:00	12,20	100,50	21,64	3,35	8,74	69,04	0,09
23:00:00	12,23	101,83	21,59	3,45	8,67	68,76	0,10

Based on Table 8, the **PM₁₀** forecasting results show a decrease from 00:00 with a value of 13.59 to reach the lowest value at 11:00 am at 11.66. After that, the **PM₁₀** value continues to increase until it reaches 12.23 at 23:00. The **PM_{2.5}** forecasting results experienced a significant decrease from 172.27 at 00:00 to 94.75 at 15:00. The **PM_{2.5}** value increased again until it reached 101.83 at 23:00. The **SO₂** forecasting results showed a consistent decrease over 24 hours, from an initial value of 22.90 to 21.59. In contrast, the **CO** forecasting results showed an increase over 24 hours, from an initial value of 1.87 to 3.45. The **O₃** forecasting result also experienced a consistent decrease over 24 hours, from 9.96 to 8.67. Likewise for **NO₂**, the forecasting results show a consistent decrease from the initial value of 76.61 to 68.76 at 23:00. Meanwhile, the **HC** forecasting results showed a slight increase from the initial value of 0 to 0.10 at 23:00.

The Air Pollution Index (ISPU) value is determined by the parameter that has the highest ISPU value at that time. Based on the forecasting results, **PM_{2.5}** is the determining parameter because it shows the highest ISPU value compared to other parameters. Based on the **PM_{2.5}** value, the air quality in Semarang City on January 1, 2025 at 00:00 to 10:00 is in the range of 101-173 which is included in the unhealthy category. In this condition, the air quality can endanger human health, especially for people with asthma and heart disease. Furthermore, from 11:00 a.m. to 21:00 p.m., the ISPU value decreases to the range of 94-100, indicating that the air quality in Semarang City is in the moderate category, meaning that the air quality is still acceptable for human, animal and plant health. However, from 22:00 to 23:00, the air quality returned to the unhealthy category as the ISPU value increased to over 100. Thus, in general, it was found that the air quality in Semarang City on January 1, 2025 was in the moderate to unhealthy category. This is in line with the results of Khusna et al. (2023) who forecasted **PM_{2.5}** concentrations in Semarang City and showed that the air quality in the region was in the unhealthy category. One of the main sources of **PM_{2.5}** in Semarang City is emissions from motor vehicles. High motor vehicle activity generates large amounts of exhaust emissions, which contribute to increased concentrations of **PM_{2.5}** and cause a decline in air quality. This condition can have adverse health effects, especially for vulnerable groups such as

people with asthma, and heart disease. Exposure to $\text{PM}_{2.5}$ can cause various health problems, such as lung cancer, respiratory infections, chronic obstructive pulmonary disease (COPD), and even increase the risk of death.

The forecast decreases smoothly during the day for most parameters. Smooth trajectories are typical of multi-step forecasts in which each prediction is used as input to the next, and they do not capture sudden pollution episodes, so the uncertainty of the later hours is larger than that of the first hours.

3.7. Discussion

The results show that the choice of the learning rate matters more than the choice of architecture: the learning rate of 0.001 gives the lowest errors for all models, and a rate of 0.1 makes the errors 9 to 17 times larger, whereas the three architectures differ by less than 0.5% in RMSE at their best settings. The best architectures also differ in size, with 100 neurons for LSTM and 50 neurons for GRU and Hybrid, which means that the Hybrid model reaches the highest accuracy with the smaller network and is the most tolerant of a wrong learning rate. These properties, and not the size of the accuracy gain, are the main practical advantage of the Hybrid LSTM-GRU in this data set.

The finding that the Hybrid model is the most accurate agrees with Yurtsever (2023) for unemployment rates and with Tripathy et al. (2023) for stock market indices, and the finding that GRU is more accurate than LSTM (MAE 0.2215 against 0.2470) agrees with Nilsen (2022) for LQ45 stock prices. The learning rate and neuron pattern agrees with Yulisa et al. (2023). In contrast to the studies of air quality in Indonesia that used ARIMA (Khusna et al., 2023), DSARIMA (Koesoemaningroem et al., 2021), or support vector regression (Rahmat et al., 2023), this study uses recurrent networks on hourly data of all seven ISPU parameters, and its forecast of an unhealthy category for $\text{PM}_{2.5}$ is in line with Khusna et al. (2023), who also found the unhealthy category for Semarang. Because the data sets, horizons, and error scales differ, the comparison is limited to the direction of the results.

The novelty of the findings is that the Hybrid LSTM-GRU is shown, for the ISPU of Semarang City, to be the most accurate and the most robust to the learning rate, but only marginally more accurate than GRU. For the authorities, the forecast that $\text{PM}_{2.5}$ determines the index and that air quality is unhealthy in the early morning and late evening can support early warning and advice for vulnerable groups, in line with the importance of controlling motor vehicle emissions. Several limitations should be noted. Each configuration appears to have been trained once, the differences between the models are small and may lie within the variation caused by random initialization, no significance test was made, the best configuration was selected on the test data, and no baseline such as a persistence forecast or ARIMA was included. Future research should repeat the training with different seeds, add a validation set and baseline models, and test deeper or bidirectional architectures and additional variables such as meteorological data.

4. Conclusion

The three research questions are answered as follows. First, the best architecture of the LSTM uses 100 neurons and a learning rate of 0.001 (mean RMSE 0.7631 and MAE 0.2470), that of the GRU uses 50 neurons and a learning rate of 0.001 (0.7600 and 0.2215), and that of the Hybrid LSTM-GRU uses 50 neurons and a learning rate of 0.001 (0.7597 and 0.2193); the learning rate of 0.001 is the best for all models. Second, the Hybrid LSTM-GRU is the most accurate method for forecasting the Air Pollution Index of Semarang City, but its advantage is small (0.04% in RMSE and 1.0% in MAE relative to GRU, and 11.2% in MAE relative to LSTM), so the three methods are close in accuracy and the Hybrid model is more clearly preferable in its robustness to the learning rate and its smaller network.

Third, the 24-hour forecast of the Hybrid model for January 1, 2025 shows $\text{PM}_{2.5}$ as the determining parameter: air quality is unhealthy from 00:00 to 10:00 and from 22:00 to 23:00 and moderate from 11:00 to 21:00. The forecast can be used for early warning, especially for vulnerable groups. Further research is recommended to repeat the training with different random seeds, to include baseline models and a validation set, and to add meteorological variables.

References

- Badan Pusat Statistik Provinsi Jawa Tengah. (2025). *Jumlah kendaraan bermotor menurut kabupaten/kota dan jenis kendaraan di Provinsi Jawa Tengah (unit)*. <https://jateng.bps.go.id/id/statistics-table/3/VjJ3NGRGa3dkRk5MTIU1bVNFOTVVbmQyVURSTVFUMDkjMyMzMzAw/jumlah-kendaraan-bermotor-menurut-kabupaten-kota-dan-jenis-kendaraan-di-provinsi-jawa-tengah--unit-.html?year=2024>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>
- Farwati, M., Salsabila, I. T., Navira, K. R., & Sutabri, T. (2023). Analisa pengaruh teknologi artificial intelligence (AI) dalam kehidupan sehari-hari. *Jursima*, 11(1), 39–45.
- Fauzan, F. D., Rayhan, D. A., Putri, H. M., & Kartiasih, F. (2024). Peramalan konsentrasi $\text{PM}_{2.5}$ menggunakan model ARCH/GARCH dan long short-term memory (studi kasus: Kota Jakarta Pusat). *Infomatek*, 26(1), 27–44.

- Health Effects Institute. (2024). *State of global air 2024: Special report*. Health Effects Institute.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- IQAir. (2025). *World's most polluted cities*. <https://www.iqair.com/world-most-polluted-cities>
- Khusna, N. F., Aulia, S., Amaria, S., Rahmah, A., Sanmas, S. A., & Fauzi, F. (2023). Peramalan kualitas udara di Semarang menggunakan metode autoregressive integrated moving average (ARIMA). *Prosiding Seminar Nasional Unimus*, 6.
- Koesoemaningroem, N., Endroyono, E., & Nugroho, S. M. S. (2021). Peramalan pencemaran udara di Kota Surabaya menggunakan metode DSARIMA dengan pendekatan percentile error bootstrap (PEB). *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 8(5).
- Ministry of Environment and Forestry. (2020). *Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor P.14/MENLHK/IKLIM/7/2020 tentang Indeks Standar Pencemar Udara*. Ministry of Environment and Forestry of the Republic of Indonesia.
- Nilsen, A. (2022). Perbandingan model RNN, model LSTM, dan model GRU dalam memprediksi harga saham-saham LQ45. *Jurnal Statistika dan Aplikasinya*, 6(1), 137–147.
- Perdana, D., & Muklason, A. (2023). Machine learning untuk peramalan kualitas indeks standar pencemar udara DKI Jakarta dengan metode hibrid ARIMAX-LSTM. *ILKOMNIKA*, 5(3), 209–222.
- Rahmat, R. W., Annas, S., & Rais, Z. (2023). Analisis support vector regression (SVR) untuk meramalkan indeks kualitas udara di Kota Makassar. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 5(3), 104–117.
- Saputra, N. W. (2023). Penerapan deep learning menggunakan gated recurrent unit untuk memprediksi harga minyak mentah dunia. *Building of Informatics, Technology and Science (BITS)*, 5(1), 86–94.
- Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A. K., & Dixit, S. (2022). Artificial intelligence technologies for forecasting air pollution and human health: A narrative review. *Sustainability*, 14(16), 9951.
- Tripathy, N., Parida, S., & Nayak, S. K. (2023). Forecasting stock market indices using gated recurrent unit (GRU) based ensemble models: LSTM-GRU. *International Journal of Computer and Communication Technology*, 9(1).
- Yulisa, P. N., Al Haris, M., & Arum, P. R. (2023). Peramalan nilai ekspor migas di Indonesia dengan model long short term memory (LSTM) dan gated recurrent unit (GRU). *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, 16(1), 328–341.
- Yurtsever, M. (2023). Unemployment rate forecasting: LSTM-GRU hybrid approach. *Journal for Labour Market Research*, 57(1), Article 18.