JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 1, January 2026, pp. 1480-1494

Analysis Comparison of BiLSTM and BiGRU Models for Aircraft Visibility Prediction

Nayla Fitriyatus Saidah[1], Nurissaidah Ulinnuha[1*], Yuniar Farida[1]
[1]Department of Mathematics, Sunan Ampel State Islamic University, Indonesia
nuris.ulinnuha@uinsa.ac.id

Article History:
Received : 11-09-2025
Revised : 11-11-2025
Accepted : 14-11-2025
Online : 01-01-2026

Keywords:
Aircraft Visibility Prediction;
BiGRU;
BiLSTM;
Deep Learning;
Time Series.

ABSTRACT

Severe weather conditions such as fog and heavy precipitation pose significant threats to aviation safety. Accurate prediction of aircraft visibility is therefore essential to support operational decision-making and reduce the likelihood of accidents. This study aims to compare and evaluate the performance of two bidirectional deep learning models, BiLSTM and BiGRU, in predicting aircraft visibility using historical meteorological data from BMKG Juanda Sidoarjo. The novelty of this research lies in applying and comparing bidirectional recurrent architectures for visibility prediction, an approach rarely explored in aviation meteorology, to assess their capability in capturing temporal dependencies within time-series visibility patterns. Both models were trained using hyperparameter tuning, with the best configuration obtained from a 24-hour input window, batch size of 32, 64 neurons, a dropout rate of 0.1, and 100–200 epochs. The dataset was divided into training and testing sets (80:20), and model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess both predictive accuracy and computational efficiency. The results indicate that while BiLSTM achieved slightly higher accuracy, BiGRU demonstrated superior overall efficiency, obtaining competitive error metrics (MSE = 1.50×10^6, RMSE = 1,223.5, MAPE = 19.35%) compared to BiLSTM (MSE = 1.58×10^6, RMSE = 1,258.1, MAPE = 19.50%). BiGRU's advantage lies in its simpler structure and faster computation, which reduce training complexity without sacrificing forecast accuracy. Overall, this research contributes to the development of efficient bidirectional time-series models for aviation meteorology, offering a practical framework for real-time visibility forecasting in computationally limited environments. The balance between accuracy, speed, and model simplicity makes BiGRU a more scalable and applicable choice for enhancing flight safety operations.

https://doi.org/10.31764/jtam.v10i1.34698

A. INTRODUCTION

Aviation safety remains one of the most critical priorities in modern air transportation, where environmental factors such as fog, heavy rain, and storms significantly affect flight operations (Wang et al., 2024). A serious threat to safety is poor weather conditions, such as dense fog, heavy rain, or storms, which can significantly reduce aircraft visibility (AlBesher & AlMusallam, 2023). This is a challenge for pilots, especially when taking off or landing, as it increases the possibility of accidents and reduces passengers' comfort and airline service confidence (Adukwu et al., 2024). Such accidents were witnessed in the media where five flights at Juanda Airport had to divert to Ngurah Rai Airport, Bali, due to poor visibility caused by bad weather (DetikCom, 2024). The worst possible visibility accepted for air traffic operation in

aviation is usually 5 to 8 km. Less than this, there is a significantly increased risk of takeoff or landing. This concurs with ICAO Visual Meteorological Conditions (VMC) minima, which specify a minimum of 5 km of flight visibility at or below 10,000 ft, and 8 km above 10,000 ft. Less than this, there is a significantly increased risk of takeoff or landing (Skybrary, n.d.). Thus, precautionary measures such as routed diversion or flight delay are usually implemented to prevent danger. These measures are implemented for the safety of passengers, and when regular visibility is restored, delayed or diverted flights are safe to land. As a result, aircraft low visibility is a critical challenge that must be adequately foreseen by all the players in the aviation industry (Akintunde et al., 2025).

Aircraft visibility prediction is an effort to minimize the negative impact of visibility caused by bad weather, which not only prevents accidents but also maintains business efficiency in the aviation industry (C.-J. Chen et al., 2023). Aircraft visibility prediction can be done using methods that utilize historical visibility data as input for past visibility observations (Singh et al., 2024). Deep learning is increasingly being used in weather research because it can handle complex, nonlinear, and dynamic data patterns more effectively than conventional approaches in this case, deep learning models such as Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) have been widely used to build time series models. These bidirectional models can learn past and future dependencies simultaneously, resulting in higher prediction efficiency compared to simple unidirectional models (Unlu & Peña, 2025).

BiGRU can learn bidirectional temporal patterns in rainfall data in Bandung with faster computation than the more complex BiLSTM. In addition, BiGRU is also reliable in solving the vanishing gradient problem Another study found that the CNN-LSTM model can outperform conventional deep learning methods such as CNN, GRU, LSTM, and BiLSTM in predicting cryptocurrency movements. However, the CNN-LSTM architecture is unidirectional, so it cannot capture temporal dependency patterns. This limitation creates a research gap that can be addressed by applying BiLSTM and BiGRU, which can learn future and past sequence patterns, thereby improving model performance (Shankar & Sahana, 2024). The conclusion demonstrates the direction in which algorithm selection should be set, depending on the type of dataset, where volatile data is best represented by BiGRU, and complex pattern stable data is best described by BiLSTM (Shaikh & Ramadass, 2024).

Although BiGRU and BiLSTM have been widely applied in various time series forecasting domains, their specific use for aircraft visibility prediction remains very limited. Moreover, few studies have directly compared the performance of these two bidirectional models in the context of aviation safety, where visibility is a crucial operational factor. This research addresses that gap by developing and evaluating both models using historical meteorological data from BMKG Juanda Sidoarjo to determine which architecture performs more effectively in predicting aircraft visibility. Scientifically, this study contributes to a deeper understanding of how bidirectional deep learning architectures, BiLSTM and BiGRU, handle nonlinear temporal patterns in meteorological time series data. By directly comparing their performance, this research provides empirical evidence on how these two models handle complex atmospheric patterns. Practically, the findings of this study are expected to serve as a foundation for

developing more reliable early-warning and decision-support systems that can enhance flight safety under low-visibility conditions.

B. METHODS

1. Preprocessing Data

The preprocessing stage begins with linear interpolation, which is applied to handle missing visibility records and ensure data continuity. This technique assumes that short-term changes in visibility follow a smooth temporal transition, which is a reasonable approximation for meteorological variables that typically vary gradually over time (Wang, 2025). Then, MinMaxScaler normalization is performed to rescale all variables into the range of 0–1. This step ensures uniform numerical magnitude among input features and prevents any single variable from dominating the training process, improving model convergence stability (Raju et al., 2020). At the time-series windowing stage, a 24-hour window was selected because meteorological factors such as temperature, humidity, and visibility generally exhibit daily cyclic patterns. Capturing one full diurnal cycle allows the model to learn dependencies between day and night visibility variations effectively (Shi et al., 2022). Finally, the dataset is divided into training and testing subsets using an 80:20 ratio. This ratio is commonly adopted in time-series forecasting to maintain a balance between model learning capacity and evaluation reliability. Cross-validation was not implemented in this study to save computational time, as it requires multiple training iterations that are computationally expensive for deep learning models. Since the dataset exhibits a continuous temporal sequence without random sampling, a single 80:20 split was deemed sufficient to represent the generalization ability of the models without iterative resampling. The research process is described in Figure 1.

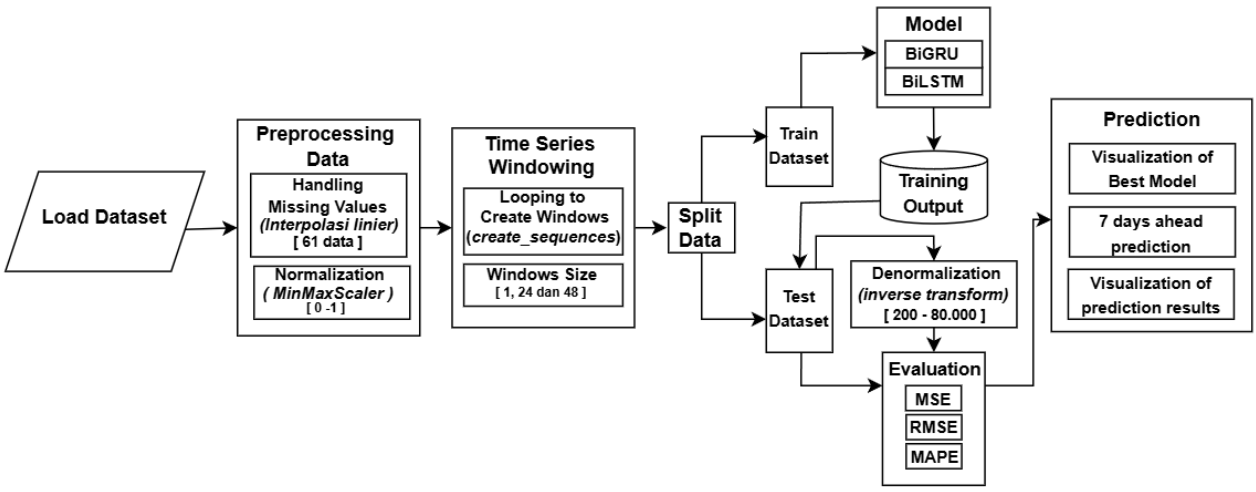

Figure 1. Research Method Flowchart

2. Bidirectional Gated Recurrent Unit (BiGRU)

The Bidirectional Gated Recurrent Unit (BiGRU) is a type of Recurrent Neural Network (RNN) designed to learn temporal dependencies from sequential data in both forward and backward directions (S. Wang et al., 2022). It consists of two GRU layers that process the same sequence in opposite directions and then combine their hidden states in the output layer. This bidirectional structure enables the model to capture both preceding and succeeding contextual information within a time sequence, resulting in a more comprehensive temporal representation (Li et al., 2023). For meteorological data such as visibility, humidity, and temperature, where patterns often repeat within daily or seasonal cycles, BiGRU is well suited to identify these dynamic relationships. Its dual-directional learning helps recognize visibility fluctuations more effectively by considering the influence of past and future conditions simultaneously. Figure 2 illustrates the BiGRU architecture used in this study, which integrates two GRU layers running in opposite directions before merging into a fully connected output layer for visibility prediction.

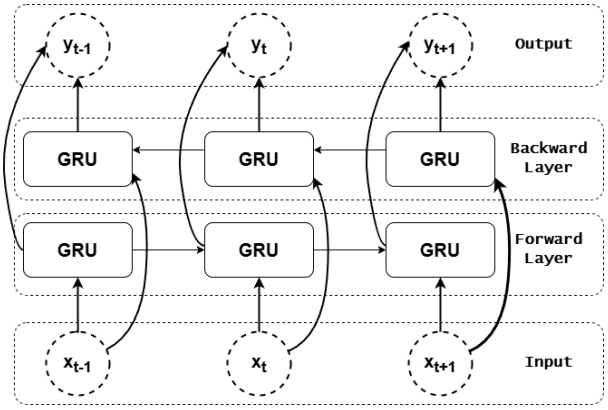

Figure 2. The Architecture of the BiGRU Model (Pranida & Kurniawardhani, 2022)

3. Bidirectional Long Short-Term Memory (BiLSTM)

The Bidirectional Long Short-Term Memory (BiLSTM) model extends the standard LSTM architecture by enabling learning in both forward and backward directions. BiLSTM employs an internal memory mechanism that helps preserve long-term dependencies while controlling the flow of information through three gates, input, forget, and output. This structure enables the model to effectively capture long-term and short-term dependencies within sequential data. (Tyas, 2024). In the context of visibility forecasting, BiLSTM provides a deeper temporal representation by analyzing both past and future meteorological trends simultaneously. This ability to capture dual-directional temporal relationships makes BiLSTM particularly powerful for modeling atmospheric dynamics that influence visibility fluctuations. As illustrated in Figure 3, BiLSTM combines two LSTM layers one reading the input sequence forward and the other backward whose outputs are merged to form the final prediction layer.

.

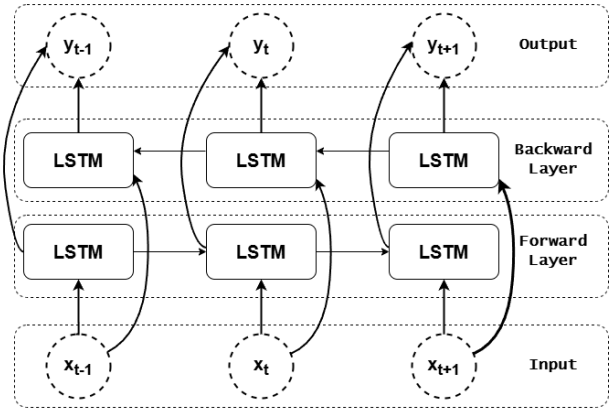

Figure 3. The Architecture of the BiLSTM Model (Pranida & Kurniawardhani, 2022)

4. Denormalization Process and Model Evaluation

The best model is attained by validation against the test dataset, and then a denormalization stage using the reverse process of the initial normalization to transform the data back to its original level, so that the prediction outcomes can be meaningfully interpreted (Qin et al., 2024). The prediction begins with an extensive testing of the selected model to forecast visibility for the next seven days. This is performed based on three critical metrics: Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These are employed because they give the complete picture of the model's performance regarding absolute error, prediction deviation, and relative error. The combination of MSE, RMSE, and MAPE enables a comparison between the magnitude of errors and the proportion of error to actual values, such that the most accurate and highest quality model is the smallest MSE, RMSE, and MAPE. These three metrics will serve as benchmarks in this study, with the following formulas (Li, 2023), as shown in Table 1.

Table 1. Evaluation Metric Formula for Prediction Models

MSE	RMSE	MAPE
$= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	$= \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{\hat{y}_i - y_i}{y_i}\right\| \times 100\%$

The interpretation of the MAPE value can be classified based on its range, as shown in Table 2, which provides a qualitative guide to the model's forecasting accuracy.

Table 2. MAPE Value Range (Montaño Moreno et al., 2013)

Range MAPE	<10%	10% - 20%	20% - 50%	>50%
Interpretation	Very Accurate Forecast	Good Forecast	Fairly Good Forecast	Inaccurate Forecast

5. Comparative Statistical Analysis of Models

To compare the performance of the BiGRU and BiLSTM models, a series of statistical tests were conducted on the prediction results obtained from both models. The analysis began with a normality test using the Shapiro–Wilk test to verify whether the error data from each model followed a normal distribution. If the test produced a significance value (Sig.) greater than 0.05, the data were considered normally distributed. Subsequently, a homogeneity of variance test was performed using Levene's test to assess whether the variances between the two models were statistically equal. The data were regarded as homogeneous if the Levene's test significance value exceeded 0.05. If both assumptions normality and homogeneity were satisfied, a paired sample t-test was then applied to determine whether there was a statistically significant difference between the performance of the BiGRU and BiLSTM models. The test was conducted at a significance level of $\alpha = 0.05$ with degrees of freedom (df) = n – 1, where n represents the number of trials. The decision criteria of t-test were defined as follows:

a. If the significance value (Sig.) < 0.05, the difference in model performance is statistically significant, indicating that one model outperforms the other.
b. If the significance value (Sig.) ≥ 0.05, the difference is not statistically significant, implying that both models perform comparably.

6. Visualization of Prediction Results

The model with the best evaluation performance is applied to predict the aircraft visibility over the next seven days, using the most recent data to ensure the relevance of the predictions. The prediction results are then visualized alongside the actual data, including training and testing data, to facilitate comparative analysis and allow for a visual assessment of the model's prediction accuracy.

C. RESULT AND DISCUSSION

1. Data and Preprocessing

The research data was collected hourly over one year, from January 1, 2024, at 01:00 to January 1, 2025, at 00:00, resulting in 8,784 data points. Each entry records the observation time and the visibility value in meters, as shown in Figure 4.

	time	(UTC)	Visibility
0	1/1/2024	1:00	5000
...	...	...	...
59	1/3/2024	12:00	5000
60	1/3/2024	13:00	NaN
61	1/3/2024	14:00	NaN
...	...	...	...
8100	12/3/2024	13:00	8000
8101	12/3/2024	14:00	NaN
8102	12/3/2024	15:00	7000
...	...	...	...
8308	12/12/2024	5:00	8000
8309	12/12/2024	6:00	NaN
8310	12/12/2024	7:00	5000
...	...	...	...
8387	12/15/2024	12:00	7000
8388	12/15/2024	13:00	NaN
8389	12/15/2024	14:00	5000
...	...	...	...
8781	12/31/2024	22:00	8000
8782	12/31/2024	23:00	8000
8783	1/1/2025	0:00	NaN

(a) Research Data

	Visibility_interpolated
60	5000
61	5000
62	5000
63	5000
64	5000
...	...
7155	10000
...	...
8101	7500
...	...
8309	6500
...	...
8388	6000
...	...
8783	8000

(b) Linear Interpolated Data

	Visibility
0	0.060150
1	0.122807
2	0.122807
3	0.122807
4	0.122807
5	0.122807
6	0.122807
...	...
8777	0.073684
8778	0.060150
8779	0.060150
8780	0.060150
8781	0.097744
8782	0.097744
8783	0.097744

(c) Normalized Data

Figure 4. Visualization of Data Preprocessing Results

The data set has 61 data points with missing values. Linear interpolation fills in missing values by drawing a straight line between the two closest data points, with the results shown in Figure 4b (J. Wang, 2025). Next, MinMaxScaler normalisation is performed, with the results presented in Figure 4c. The results of time series windowing with 24 inputs are shown in Figure 5a, and the results of 80:20 data division are shown in Figure 5b. Next, the data from windowing and the division of the training and testing sets are formulated as model inputs for the modeling stage.

Sample Window Input :
2024-01-01 00:00:00 → 2024-01-02 00:00:00

	Time	X_norm	X_original
24	2024-01-01	0.060150	5000.0
23	2024-01-01	0.122807	10000.0
22	2024-01-01	0.122807	10000.0
21	2024-01-01	0.122807	10000.0
20	2024-01-01	0.122807	10000.0
19	2024-01-01	0.122807	10000.0
18	2024-01-01	0.122807	10000.0
...	...	...	...
8	2024-01-01	0.060150	5000.0
7	2024-01-01	0.060150	5000.0
6	2024-01-01	0.035088	3000.0
8	2024-01-01	0.060150	5000.0
5	2024-01-01	0.022556	2000.0
4	2024-01-01	0.060150	5000.0
3	2024-01-01	0.060150	5000.0
2	2024-01-01	0.060150	5000.0
1	2024-01-02	0.085213	7000.0

(a)

X(t-24)	X(t-23)	X(t-22)	...	X(t-3)	X(t-2)	X(t-1)	Y(t)
0.060150	0.122807	0.122807	...	0.060150	0.060150	0.085213	0.122807
0.122807	0.122807	0.122807	...	0.060150	0.085213	0.122807	0.122807
0.122807	0.122807	0.122807	...	0.085213	0.122807	0.122807	0.122807
0.122807	0.122807	0.122807	...	0.122807	0.122807	0.122807	0.122807
0.122807	0.122807	0.122807	...	0.122807	0.122807	0.122807	0.122807

(b)

Figure 5. (a) Time Series Windowing; and (b) 80:20 Data Division

2. Model Configuration and Evaluation

a. Training Model

Both models, the BiGRU and BiLSTM, process data from the forward and backward directions. Each model uses a single bidirectional layer with 64 units and a dropout rate of 0.1 to reduce overfitting, followed by a dense layer with 32 neurons using the ReLU activation function. Training is carried out with batch sizes of 32 and 64, optimized using the Adam optimizer with a learning rate of 0.001, and monitored with EarlyStopping (patience = 10). Model performance is evaluated across different epochs (50, 100, 150, and 200) and historical data windows of 1, 24, and 48 hours to analyze how temporal context influences visibility prediction.

b. Denormalization Process

After obtaining the best model through testing on the testing data, the denormalization process is performed using the inverse method of the initial normalization to return the data to its original scale, allowing the predicted results to be interpreted in their actual context (Qin et al., 2024). The denormalized output data is shown at Table 3.

Table 3. Denormalized Actual Y Data

Date	Denormalized Data
1/1/2024	5000
1/1/2024	10000
1/1/2024	10000
...	...
12/31/2024	8000
12/31/2024	8000
1/1/2025	8000

c. Implementation of the BiGRU Model

After denormalization, the model is trained with the test data based on evaluation metrics set as MSE, RMSE, and MAPE for each trial during the epoch number. The best model is selected by employing the minimum MSE, RMSE, and MAPE values obtained from all experiments, and the result is used to store the best model as well as the scaler used during training. In this study, the visibility of aircraft will be accurately predicted, and it will be determined whether BiGRU or BiLSTM is the best approach. Before obtaining the optimal results for evaluation, there was a hyperparameter tuning process, as evident in Table 4.

Table 4. Hyperparameter Tuning Results for BiGRU Model

D	B	N	0	Dropout 0.1			
				MSE	RMSE	MAPE	Time
1	32	64	50	1.77×10^6	1330.4	19.998	31s
			100	1.78×10^6	1334.9	20.65	26s
			150	1.78×10^6	1333	20.064	30s
			200	1.77×10^6	1331	20.484	35s
24	32	64	50	1.58×10^6	1255.5	19.907	258s
			100	1.50×10^6	1223.5	19.346	265s
			150	1.61×10^6	1267.2	20.052	283s

D	B	N	0	Dropout 0.1			
				MSE	RMSE	MAPE	Time
			200	1.59 x 10^6	1259.1	19.459	273s
48	32	64	50	1.89 x 10^6	1375	21.077	136s
			100	1.88 x 10^6	1369.4	21.273	145s
			150	1.90 x 10^6	1380.1	21.106	176s
			200	1.88 x 10^6	1372.3	21.292	183s

The training process showed that increasing Historical Data, Batch Size, Neuron, and Epoch values, and decreasing the Dropout value, did not necessarily improve performance or cut down computation. Therefore, it is necessary to choose the proper set of parameters. The best combination of Historical Data (D) time series windowing for 1 hour, 24 hours, and 48 hours with other parameters such as Batch Size (B), Neuron (N), Epoch (O), and Dropout from the grey-lined row (D=24, B=32, N=64, O=100, Dropout=0.1) provided evaluation metrics of MSE 1.50 x 10^6, RMSE 1223.5, and MAPE 19.346 with a computation efficiency of 265s. This is the optimal compromise between prediction accuracy and computational complexity and, therefore, is the optimal setting for calibration of the BiGRU model.

d. Implementation of the BiLSTM Model

The BiLSTM model was also executed to verify how the performance of BiLSTM compares with the BiGRU model with the same batch of hyperparameters. The results of hyperparameter tuning of the BiLSTM model are presented in Table 5, along with adjustments in time series windowing configuration of Historical Data (D) for 1 h, 24 h, and 48 h, among other parameters such as Batch Size (B), Neuron (N), Epoch (O), and Dropout, as shown in Table 5.

Table 5. Hyperparameter Tuning Results for BiLSTM Model

D	B	N	0	Dropout 0.1			
				MSE	RMSE	MAPE	Time
1	32	64	50	1.85 x 10^6	1360.5	20.295	23s
			100	1.80 x 10^6	1340.6	20.602	50s
			150	1.84 x 10^6	1358.1	20.335	20s
			200	1.79 x 10^6	1339.7	20.889	42s
24	32	64	50	1.60 x 10^6	1264.5	19.655	294s
			100	1.58 x 10^6	1258.1	19.705	196s
			150	1.59 x 10^6	1262.5	19.81	234s
			200	1.58 x 10^6	1258.1	19.504	264s
48	32	64	50	1.98 x 10^6	1408.5	22.466	154s
			100	1.99 x 10^6	1410.2	21.58	289s
			150	1.82 x 10^6	1348.8	21.282	189s
			200	1.95 x 10^6	1396.8	21.548	205s

A lesser Dropout learning process proved that a greater Batch Size, Neuron, and Epoch, along with a lower dropout rate, were not necessarily the best model. The optimal parameters and prediction objective were the best approach. The configuration beneath the grey-colored row (D=24, B=32, N=64, O=200, Dropout=0.1) yielded values of MSE 1.58 × 10^6, RMSE 1,258.1, and MAPE 19.504, with an excellent computation time of

264 seconds. The configuration offers an excellent balance between prediction performance and computational time, making it the optimal choice for hyperparameter tuning of the BiLSTM model.

e. Comparison of BiGRU and BiLSTM Models

The best results of the hyperparameter tuning for the BiGRU and BiLSTM models are shown in the following Table 4 and Table 5.

Table 6. Best Model Comparison Results

D	B	N	0	Dropout 0.1				
				Model	MSE	RMSE	MAPE	TIME
24	32	64	50	BiGRU	1.58	1255.5	19.907	258s
			50	BiLSTM	1.60	1264.5	19.655	294s
			100	BiGRU	1.50	1223.5	19.346	265s
			100	BiLSTM	1.58	1258.1	19.705	196s
			150	BiGRU	1.61	1267.2	20.052	283s
			150	BiLSTM	1.59	1262.5	19.810	234s
			200	BiGRU	1.59	1259.1	19.459	273s
			200	BiLSTM	1.58	1258.1	19.504	264s

In the hyperparameter tuning experiment with the following settings (Historical Data (D) = 24, Batch Size (B) = 32, Neuron (N) = 64, Dropout = 0.1), the BiGRU and BiLSTM models exhibited their optimal performance under different epoch settings. The BiGRU model yielded an MSE of 1.50 x 10^6, an RMSE of 1223.5, and a MAPE of 19.346, with a computation time of 265s at 100 epochs. The BiLSTM model, however, yielded an MSE of 1.58 x 10^6, an RMSE of 1258.1, and a MAPE of 19.504, with a computation time of 264s at 200 epochs. Comparing these results, BiGRU outperformed BiLSTM in all the evaluation measures, providing lower MSE, RMSE, and MAPE values while consuming nearly the same computation time. Both models' MAPE values fall in the Good Forecasting class, although BiGRU is more precise and computationally efficient in predicting visibility.

f. Comparative statistical analysis of models

Based on the results of twelve experimental trials, the BiGRU model achieved a mean value of 1319.2833 with a standard deviation of 54.2685. In contrast, the BiLSTM model obtained a higher mean of 1333.8667 and a standard deviation of 59.0397. The smaller mean value of BiGRU indicates that it produced lower prediction errors compared to BiLSTM. Therefore, from a descriptive statistical perspective, BiGRU demonstrates better predictive performance and greater stability.

The Shapiro–Wilk normality test results for the RMSE values of both models show significance values of 0.085 for BiGRU and 0.054 for BiLSTM. Both significance values are around the threshold of 0.05, so it can be concluded that the data is normally distributed and meets the normality assumption. This indicates that the residuals from both models were usually distributed. Furthermore, the Levene's test for homogeneity of variance yielded a significance value of 0.771, which is greater than 0.05, confirming that the two datasets have homogeneous variances.

Since both assumptions of normality and homogeneity were satisfied, a paired sample t-test was performed to assess whether there was a significant difference between the two models. The test produced a t-statistic value of −2.428 and a t-critical value of 2.201 at a 5% significance level, with a p-value of 0.034. Because $|t$ *hitung*| > t table and $p < 0.05$, the null hypothesis (H_0) is rejected, indicating a statistically significant difference between the BiGRU and BiLSTM models. In conclusion, the BiGRU model achieved lower average prediction errors and more consistent results, confirming that it outperforms BiLSTM in terms of predictive accuracy and model stability. The results of the normality, homogeneity, and paired t-tests are summarized in Table 7.

Table 7. Statistical Analysis for Comparing Model Performance

Type of Test	Sig. (p-value)	Conclusion
Normality Test (Shapiro–Wilk)	BiGRU = 0.085 BiLSTM= 0.054	The data are normally distributed
Homogeneity Test (Levene's Test)	0.771	The variances of both models are homogeneous
Paired Sample t-Test	0.035	There is a significant difference; BiGRU performs better

The comparison plot also verifies this outcome, with the BiGRU prediction line closely following the actual trend of visibility, particularly in sequences with oscillating values. Although small fluctuations occur in highly low visibility values, BiGRU exhibits closer agreement with real data than BiLSTM. All required outcomes thus confirm that BiGRU is more accurate and computationally less expensive, making it the best suited model for aircraft visibility prediction in this study, as shown in Figure 6.

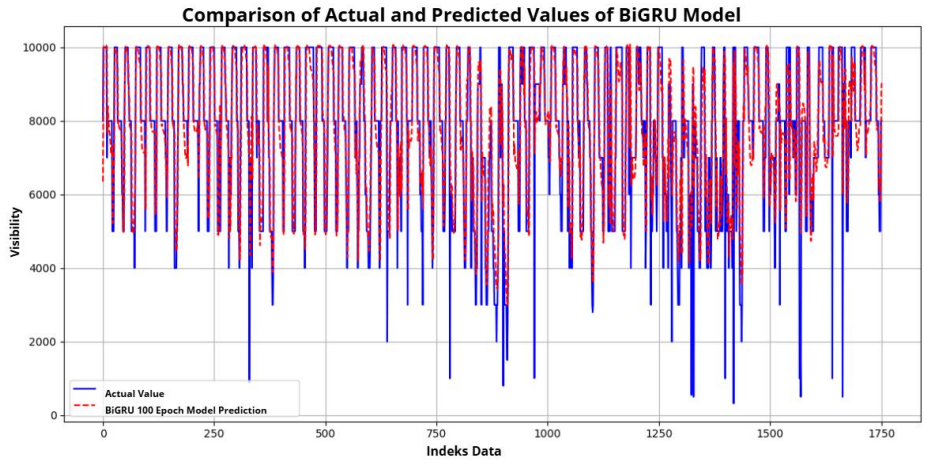

Figure 6. Actual and Predicted Values of the Best Model

To further validate the results of this study, a comparison was conducted with previous research related to airport visibility prediction in Indonesia. Both studies applied machine learning approaches using meteorological parameters. However, they differ in modelling techniques and data sources. The comparison aims to

highlight how the present study aligns with or improves upon earlier works in terms of prediction accuracy and methodological approach, as shown in Table 8.

Table 8. Comparison of Previous Studies on Aircraft Visibility Prediction

Previous Research	Location	Method	RMSE
(Moonlight et al., 2023)	Juanda Airport	Artificial Neural Network (Backpropagation)	677.6
(Kharisma et al., 2025)	Soekarno-Hatta Airport	Random Forest Regressor	772.3
This study	Juanda Airport	BiGRU	1,223.5

Based on the comparison, the Backpropagation Neural Network used by Moonlight et al. (2023) at Juanda Airport achieved an RMSE of 677.6, while the Random Forest Regressor developed by Kharisma et al. (2025) for Soekarno–Hatta Airport obtained an RMSE of 772.3. Despite differences in data scale and evaluation units, both models demonstrated strong predictive performance in visibility forecasting. This comparison reinforces that applying data-driven machine learning models can significantly enhance aviation meteorology. The present study's BiGRU model adopts a bidirectional deep learning approach for visibility prediction at Juanda Airport. Although the RMSE value (1,223.5) is higher than those obtained in previous studies, the model demonstrates the applicability of deep sequential architectures in capturing temporal patterns in meteorological data and provides a basis for further model refinement.

g. Prediction Process with the Best Model

The next step involves using the best model to predict 168 data points, equivalent to 7 days, from January 1, 2025, at 00:00 to January 7, 2025, at 23:00. The prediction results are shown in Figure 6.

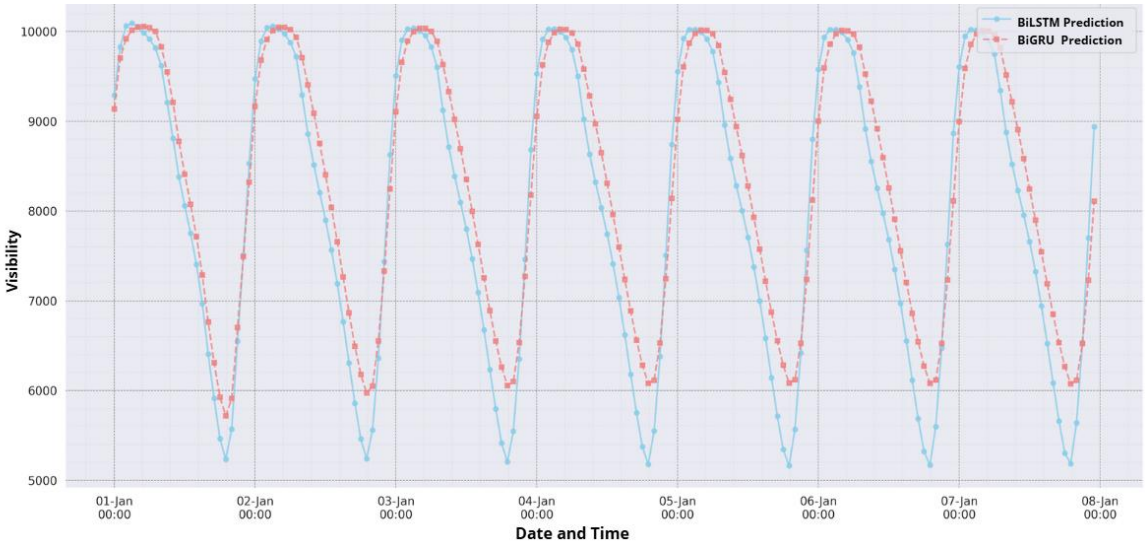

Figure 6. Prediction Results

As can be observed in Figure 6, there are noticeable dissimilarities in the patterns, ranges, and temporal responses of the BiLSTM and BiGRU model predictions. Overall, both models follow the diel cycle of visibility, with higher values during the daytime and steep decreases at nighttime. The BiGRU model predicts visibility in the range of approximately 6000–10000 meters, while the BiLSTM model predicts a comparatively wider range of approximately 5200–10000 meters, indicating BiLSTM accounts for more extreme variations. BiLSTM is also sensitive to short-term dynamics, particularly during the morning transition, when visibility rises abruptly due to surface warming and fog evaporation.

At night, BiLSTM showed a sharper decline in visibility than BiGRU, in line with findings related to the effects of humidity and fog (Boutle et al., 2022). The most significant difference occurred on 1 January 2025 at 03:00, when BiLSTM predicted a decline of more than 120 metres compared to BiGRU. During the day, both models predicted visibility above 9,000 metres, and during the seven-day observation period, most of the predictions exceeded 6,000 metres, meeting international safety standards. BiLSTM and BiGRU demonstrate more accurate model performance with MAPE of 19.50% and 19.35%, respectively, compared to the RNN and U^{2}Net models (He et al., 2024) and (J. Chen et al., 2023).

BiGRU excels in its simpler structure, which utilizes only two gates (reset and update), compared to BiLSTM, which employs three gates (input, forget, and output). The simpler structure reduces computational overhead, lowers training time, and avoids the phenomenon of overfitting without affecting the capacity to learn temporal dependencies efficiently. The efficiency of BiGRU renders it more suitable for real-time or large-scale prediction tasks, where resource optimization and speed are essential. While BiLSTM can grasp complex temporal patterns, BiGRU achieves nearly the same, if not better, predictive performance with lower computational complexity and less complex structure. Overall, BiGRU is superior due to its stability, lower error, and lighter computational requirements, making it more efficient for aircraft visibility prediction.

D. CONCLUSION AND SUGGESTIONS

Both BiGRU and BiLSTM models demonstrated reliable performance in predicting aircraft visibility, with MAPE values around 19%. However, BiGRU achieved better results with lower error metrics (MSE = 1.50×10^6; RMSE = 1,223.5; MAPE = 19.34%) compared to BiLSTM (MSE = 1.58×10^6; RMSE = 1,258.1; MAPE = 19.50%), while also requiring less computational time. The superior performance of BiGRU can be attributed to its bidirectional recurrent architecture and simplified gating mechanism, which consists of only two gates: reset and update. This design allows the model to efficiently capture temporal dependencies from both past and future sequences while reducing computational overhead and minimizing the vanishing gradient problem. These structural advantages enable BiGRU to produce accurate predictions with greater stability and faster convergence compared to BiLSTM. From a practical standpoint, this finding holds significant importance for aviation safety. Accurate and timely visibility forecasting enables air traffic controllers and pilots to make informed operational decisions, thereby minimizing the risks of delays, diversions, and potential flight accidents under low-

visibility conditions. From a scientific perspective, the study contributes to meteorological forecasting research by providing empirical evidence of the effectiveness of bidirectional recurrent networks, particularly BiGRU.

For future research, it is recommended to move beyond parameter tuning and explore broader and more impactful extensions. These include developing multivariate forecasting models that integrate additional meteorological variables such as humidity, temperature, and wind speed to enhance predictive robustness. Moreover, the application of time series cross-validation is strongly encouraged to improve the reliability of model evaluation and ensure a more accurate representation of temporal dependencies in sequential data. Implementing this approach would prevent overfitting and yield more generalizable forecasting results.

ACKNOWLEDGEMENT

The author would like to express sincere gratitude to Universitas Islam Negeri Sunan Ampel Surabaya, Faculty of Science and Technology, Department of Mathematics, and BMKG Juanda for the support in completing and publishing this study.

REFERENCES

Adukwu, S., James, T. O., & Onwuka, G. I. (2024). Effects of Human Factors and Some Selected Weather Variables on Plane Crash. *International Journal of Science for Global Sustainability*, *10*(2), 190–196. https://doi.org/10.57233/ijsgs.v10i2.665

Akintunde, E. A., Fada, S. J., Adamu, U. M., Goyol, S. S., Bombom, L. S., & Nyango, K. C. (2025). Weather-Induced Flight Disruptions: Analysing Trends, Implications and Mitigation Strategies at the Jos Airport, Plateau State, Nigeria. *African Journal of Geographical Sciences*, *5*, 1 & 2. https://doi.org/10.5281/zenodo.15010571

AlBesher, Q., & AlMusallam, S. (2023). Examining the role of weather in aircraft accidents. *International Journal of Innovative Science and Research Technology, 8*(12), Article 705. https://doi.org/10.5281/zenodo.10405099

Boutle, I., Angevine, W., Bao, J.-W., Bergot, T., Bhattacharya, R., Bott, A., Ducongé, L., Forbes, R., Goecke, T., Grell, E., Hill, A., Igel, A. L., Kudzotsa, I., Lac, C., Maronga, B., Romakkaniemi, S., Schmidli, J., Schwenkel, J., Steeneveld, G.-J., & Vié, B. (2022). Demistify: A large-eddy simulation (LES) and single-column model (SCM) intercomparison of radiation fog. *Atmospheric Chemistry and Physics*, *22*(1), 319–333. https://doi.org/10.5194/acp-22-319-2022

Chen, C.-J., Huang, C.-N., & Yang, S.-M. (2023). Aviation visibility forecasting by integrating Convolutional Neural Network and long short-term memory network. *Journal of Intelligent & Fuzzy Systems*, *45*(3), 5007–5020. https://doi.org/10.3233/JIFS-230483

Chen, J., Yan, M., Qureshi, M. R. H., & Geng, K. (2023). Estimating the visibility in foggy weather based on meteorological and video data: A Recurrent Neural Network approach. *IET Signal Processing*, *17*(1). https://doi.org/10.1049/sil2.12164

DetikCom. (2024, February 18). *5 pesawat mendarat di Bandara Juanda sempat dialihkan gegara cuaca buruk* [DetikJatim]. Detik. https://www.detik.com/jatim/berita/d-7150488/5-pesawat-mendarat-di-bandara-juanda-sempat-dialihkan-gegara-cuaca-buruk

He, D., Wang, Y., Tang, Y., Kong, D., Yang, J., Zhou, W., Li, H., & Wang, F. (2024). Improvement in the Forecasting of Low Visibility over Guizhou, China, Based on a Multi-Variable Deep Learning Model. *Atmosphere*, *15*(7), 752. https://doi.org/10.3390/atmos15070752

Kharisma, A., Fadhillah, M., & Haryanto, Y. D. (2025). Advancing aviation meteorology: Airport visibility prediction using random forest regressor on integrated METAR parameters. *Jurnal Ilmu dan Inovasi Fisika, 9*(2), 62–72. https://doi.org/10.24198/jiif.v9i2.65464

Li, X., & Zhang, X. (2023). A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. *Environmental Science and Pollution Research*, *30*(55), 117485-117502. https://doi.org/10.1007/s11356-023-30428-5

Li, X., Zhang, Y., Jin, J., Sun, F., Li, N., & Liang, S. (2023). A model of integrating convolution and BiGRU dual-channel mechanism for Chinese medical text classifications. *PLOS ONE*, *18*(3), e0282824. https://doi.org/10.1371/journal.pone.0282824

Montaño Moreno, J., Palmer Pol, A., Sesé Abad, A., & Cajal Blasco, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, *4*(25), 500–506. https://doi.org/10.7334/psicothema2013.23

Moonlight, L. S., Harianto, B. B., Musadek, A., Sukma, M. M., & Arifianto, T. (2023). Airport Visibility Prediction System to Improve Aviation Safety. In B. Bagus Harianto, R. Mahmud, A. Anas Arifin, E. Subagyo, A. Mu'ti Sazali, & Y. Chrisnawati (Eds.), *Proceedings of the International Conference on Advance Transportation, Engineering, and Applied Science (ICATEAS 2022)* (pp. 199–210). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-092-3_18

Pranida, S. Z., & Kurniawardhani, A. (2022). Sentiment Analysis of Expedition Customer Satisfaction using BiGRU and BiLSTM. *Indonesian Journal of Artificial Intelligence and Data Mining*, *5*(1), 44. https://doi.org/10.24014/ijaidm.v5i1.17361

Qin, D., Li, Y., Chen, W., Zhu, Z., Wen, Q., Sun, L., Pinson, P., & Wang, Y. (2024). *Evolving Multi-Scale Normalization for Time Series Forecasting under Distribution Shifts* (arXiv:2409.19718). arXiv. https://doi.org/10.48550/arXiv.2409.19718

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Shaikh, Z. M., & Ramadass, S. (2024). Unveiling deep learning powers: LSTM, BiLSTM, GRU, BiGRU, RNN comparison. *Indonesian Journal of Electrical Engineering and Computer Science*, *35*(1), 263. https://doi.org/10.11591/ijeecs.v35.i1.pp263-273

Shankar, A., & Sahana, B. C. (2024). Efficient prediction of runway visual range by using a hybrid CNN-LSTM network architecture for aviation services. *Theoretical and Applied Climatology*, *155*(3), 2215–2232. https://doi.org/10.1007/s00704-023-04751-3

Shi, J., Jain, M., & Narasimhan, G. (2022). *Time Series Forecasting (TSF) Using Various Deep Learning Models* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2204.11115

Singh, S. P., Shankar, A., & Sahana, B. C. (2024). Prediction of Low-Visibility Events by Integrating the Potential of Persistence and Machine Learning for Aviation Services. *MAUSAM*, *75*(4), 977–992. https://doi.org/10.54302/mausam.v75i4.6624

Skybrary. (n.d.). *Visual Meteorological Conditions (VMC) | SKYbrary Aviation Safety*. Retrieved September 2, 2025, from https://skybrary.aero/articles/visual-meteorological-conditions-vmc

Unlu, A., & Peña, M. (2025). Comparative Analysis of Hybrid Deep Learning Models for Electricity Load Forecasting During Extreme Weather. *Energies*, *18*(12), 3068. https://doi.org/10.3390/en18123068

Wang, F., Yuan, J., Liu, X., Wang, P., Xu, M., Li, X., & Li, H. (2024). Impacts of Flight Operations on the Risk of Runway Excursions. *Applied Sciences*, *14*(3), 975. https://doi.org/10.3390/app14030975

Wang, J. (2025). Data interpolation methods with the UNet-based model for weather forecast. *International Journal of Data Science and Analytics*, *20*(3), 2525–2538. https://doi.org/10.1007/s41060-024-00611-z

Wang, S., Shao, C., Zhang, J., Zheng, Y., & Meng, M. (2022). Traffic flow prediction using bi-directional gated recurrent unit method. *Urban Informatics*, *1*(1), 16. https://doi.org/10.1007/s44212-022-00015-z