International Journal Software Engineering and Computer Science (IJSECS)
6 (2), 2026, 496-509
Published Online August 2026 in IJSECS (http://www.journal.lembagakita.com/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i2.7332.

RESEARCH ARTICLE Open Access

Analysis of Multilayer Perceptron, Long Short-Term Memory, and Temporal Convolutional Network Modeling Algorithms for Rainfall Prediction at Soekarno–Hatta Airport

Eko Widyantoro HS [1*], **Sajarwo Anggai** [2], **Sudarno Wiharjo** [3].

[1*,2,3] Master of Informatics Engineering Program, Pamulang University, Banten, Indonesia.

**Corresponding author*: ekoarekwidyantoro1972@gmail.com.

Received: May 30, 2026; Accepted: July 5, 2026; Published: August 1, 2026.

Abstract: Tropical atmospheric variability poses challenges for daily precipitation prediction at Soekarno–Hatta International Airport, with implications for aviation safety. This study aimed to compare the performance of three deep learning architectures—Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Temporal Convolutional Network (TCN)—for daily precipitation prediction using meteorological observations from the BMKG Soekarno–Hatta Station for the period 2019–2025. A quantitative experimental approach was employed using a time-series split scheme to preserve the temporal structure of the data and reduce the risk of data leakage. Air temperature, relative humidity, atmospheric pressure, and wind speed were used as input variables, while precipitation was used as the target variable. Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results showed that TCN achieved the best predictive performance, with an RMSE of 10.32 mm, MAE of 7.15 mm, MAPE of 18.27%, and R^2 of 0.87, followed by LSTM and MLP. Permutation-based sensitivity analysis identified relative humidity and air temperature as the two most influential variables for model prediction. Overall, TCN demonstrated stronger performance in capturing temporal patterns in daily precipitation than LSTM and MLP and showed potential for future application in operational precipitation forecasting and extreme-weather early warning systems at airport environments.

Keywords: Climate Variability; Deep Learning; Mlp; Lstm; Precipitation Prediction; Tcn; Time-Series Split.

1. Introduction

Climate change over recent decades has been associated with changes in the frequency and intensity of extreme weather events, affecting strategic sectors such as air transportation (Akhtar, 2020). Weather forecasting has become increasingly challenging in tropical regions because atmospheric conditions are influenced by monsoon circulation, ocean–atmosphere interactions, and local convective processes (Ding *et al.*, 2024; Lee *et al.*, 2023). For aviation operations, accurate precipitation forecasting is important for supporting safe takeoff, landing, and en-route operations, as well as maintaining the efficiency and reliability of flight schedules (de Jesús Machado-Guillén *et al.*, 2024). Soekarno–Hatta International Airport presents particular challenges for daily weather prediction because of its location near the Java Sea and within the Jakarta metropolitan area. These geographical characteristics are associated with local atmospheric processes, including urban heat island effects, sea–land breeze circulation, and convective instability related to surface heating. Conventional forecasting systems based on Numerical Weather Prediction (NWP) can represent large-scale atmospheric conditions, but the spatial and temporal resolution of a given NWP system may limit its ability to represent local-scale meteorological processes around the airport (Chang *et al.*, 2021). This limitation creates an opportunity to examine data-driven approaches that can learn relationships between local meteorological observations and precipitation.

Recent developments in machine learning and deep learning have provided alternative approaches for modeling nonlinear meteorological time series. Multilayer Perceptron (MLP) can model nonlinear relationships between input variables, whereas Long Short-Term Memory (LSTM), as a recurrent neural network architecture, is designed to learn temporal dependencies through its memory cells and gating mechanisms (Jiang *et al.*, 2024). Temporal Convolutional Network (TCN) uses causal and dilated convolutions to capture temporal patterns over different time scales while maintaining stable gradient propagation (Bai *et al.*, 2018). These characteristics make MLP, LSTM, and TCN suitable candidates for comparative evaluation in daily precipitation prediction.

Previous studies have investigated various machine learning approaches for weather forecasting. Artificial Neural Networks have been applied to temperature nowcasting and meteorological prediction (Hudnurkar *et al.*, 2015; Meheretie *et al.*, 2021), while Random Forest has been used for weather prediction in operational settings (de Jesús Machado-Guillén *et al.*, 2024). LSTM and Generative Adversarial Networks have also been applied to improve precipitation forecasts derived from Numerical Weather Prediction systems (Jeong & Yi, 2023; Nikam & Mane, 2022). Other studies have examined deep learning approaches for rainfall and extreme precipitation prediction, including sequential LSTM models and hybrid neural network architectures (Chamatidis *et al.*, 2023; Jiang *et al.*, 2024). In Indonesia, Nugraha *et al.* (2023) investigated weather forecasting using LSTM, while Maheng *et al.* (2023) analyzed historical rainfall and temperature patterns in Jakarta. However, comparative evaluations of different deep learning architectures for daily precipitation prediction using airport-based meteorological observations in tropical Indonesia remain limited.

This study addresses this limitation by comparing three deep learning architectures—MLP, LSTM, and TCN—for daily precipitation prediction using meteorological observations from the BMKG Meteorological Station at Soekarno–Hatta Airport. A time-series split scheme is applied to preserve the chronological structure of the data and reduce the risk of information leakage between training and evaluation periods. The study also applies permutation-based sensitivity analysis to examine the relative contribution of the meteorological input variables and evaluates model performance under different seasonal conditions and extreme precipitation events.

The objectives of this study are to develop and evaluate MLP, LSTM, and TCN models for daily precipitation prediction, examine the sensitivity of meteorological variables to model predictions, and compare model performance using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The comparison is intended to identify the best-performing architecture for the study setting. The findings are expected to contribute to the application of deep learning for tropical meteorological time-series prediction and provide a basis for further evaluation of data-driven precipitation forecasting approaches in airport environments.

2. Related Work

Previous studies have explored a range of machine learning and deep learning approaches for weather forecasting, particularly for short-term prediction of precipitation and temperature. Artificial Neural Networks (ANNs) have been applied to temperature nowcasting and meteorological modeling, demonstrating their ability to represent nonlinear relationships in atmospheric data (Hudnurkar *et al.*, 2015; Meheretie *et al.*, 2021). In operational airport environments, ensemble-based methods such as Random Forest have also been applied to heterogeneous meteorological datasets to support weather prediction and operational decision-making (de Jesús Machado-Guillén *et al.*, 2024). More recent studies have applied deep learning architectures to improve Numerical Weather Prediction (NWP) outputs. LSTM and Generative Adversarial Networks (GANs), for example, have been used in precipitation forecasting and NWP post-processing to improve forecast quality and address systematic errors (Jeong & Yi, 2023; Nugraha *et al.*, 2023).

Recent literature has also emphasized the use of temporal deep learning architectures for meteorological time-series modeling, particularly when the target variable exhibits nonlinear and time-dependent behavior. Chamatidis *et al.* (2023) investigated deep learning approaches for meteorological prediction and demonstrated the potential of sequential models to capture temporal atmospheric variability. TCN has received attention in time-series forecasting because causal and dilated convolutions can represent temporal dependencies across different time scales while maintaining a parallel convolutional structure (Bai *et al.*, 2018). Chang *et al.* (2021) introduced a Markov Conditional Forward model for extreme-weather forecasting, demonstrating the application of advanced sequence-based methods to rapidly changing atmospheric conditions. These studies indicate that deep learning architectures can provide useful alternatives to conventional statistical approaches for complex meteorological time series.

Despite these developments, the application of deep learning to tropical precipitation forecasting remains an area that requires further investigation. Tropical precipitation is strongly influenced by local convection, moisture conditions, and interactions between atmospheric and surface processes, which can produce substantial temporal variability. Jiang *et al.* (2024) examined deep learning approaches for weather-related prediction, while Nugraha *et al.* (2023) investigated weather forecasting in Indonesia using LSTM. Maheng *et al.* (2023) focused on historical rainfall patterns in the Jakarta region. However, direct comparisons among different deep learning architectures using airport-based meteorological observations in tropical Indonesia remain limited. In particular, further evaluation is needed to determine how different architectures perform under seasonal variability and during periods of extreme precipitation.

The present study therefore compares MLP, LSTM, and TCN for daily precipitation prediction at Soekarno–Hatta International Airport. The comparison uses a time-series split scheme to preserve the chronological structure of the observations and reduce the risk of information leakage between training and evaluation periods. In addition, permutation-based sensitivity analysis is applied to examine the relative contribution of the meteorological input variables, while model performance is evaluated across seasonal periods and extreme precipitation events. This design provides a direct comparison of the three architectures under the same study setting and evaluation criteria.

3. Methodology

The research procedure consisted of several stages, including meteorological data collection, data preprocessing, time-series transformation, model development, hyperparameter tuning, model evaluation, and sensitivity analysis. Meteorological observations were collected from BMKG for the 2019–2025 period. Data preprocessing included missing-value treatment, outlier detection, normalization, and conversion of the time series into supervised learning sequences using a sliding-window approach. The dataset was divided into training, validation, and testing subsets using a time-series split scheme to preserve the chronological structure of the observations and reduce the risk of temporal data leakage. Three deep learning architectures—Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Temporal Convolutional Network (TCN)—were then developed and compared for daily precipitation prediction. Model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Training and validation loss curves were also examined to assess convergence and identify potential overfitting. The final analysis included variable sensitivity assessment and evaluation of model performance during extreme precipitation events to identify the best-performing architecture for the study setting.

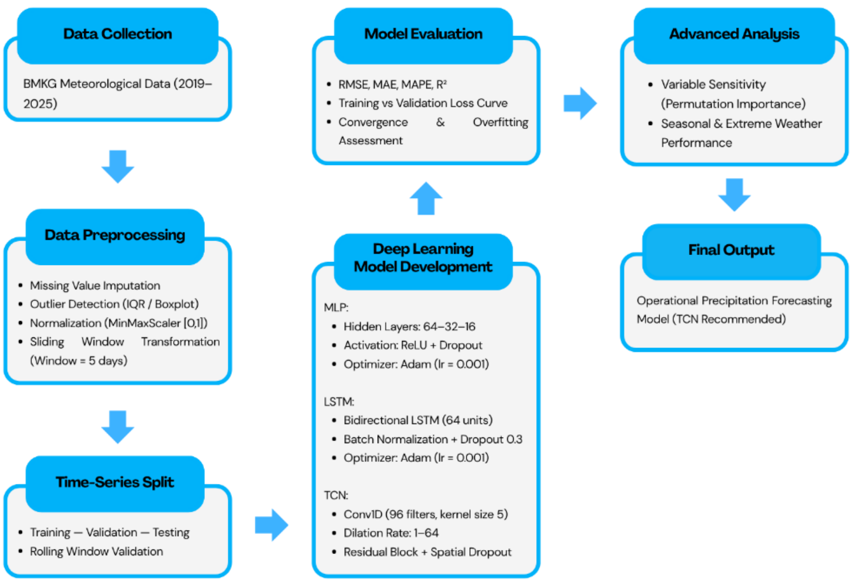

Figure 1. Research Methodology Diagram

3.1 Data

Meteorological observations were obtained from the BMKG Meteorological Station at Soekarno–Hatta International Airport, located at 6.12° S and 106.65° E. The original dataset had a temporal resolution of 10 minutes and was aggregated into daily values for daily precipitation prediction. The dataset covered the period from 2019 to 2025, yielding 2,557 daily observations. The variables included air temperature (°C), relative humidity (%), air pressure (hPa), wind speed (m/s), and daily precipitation (mm) as the target variable. The aggregation procedure for each meteorological variable should be specified, including the statistical operation used to derive daily temperature, relative humidity, air pressure, wind speed, and precipitation values.

3.2 Data Preprocessing

Data preprocessing was performed to ensure data quality and consistency before model training. Missing values were addressed using linear interpolation for gaps below 5%, while outliers were detected using the interquartile range (IQR) method. Precipitation outliers were retained because extreme precipitation values are relevant to the prediction objective. All variables were normalized to the [0,1] range using MinMaxScaler. The fitting procedure for MinMaxScaler should be specified to confirm that the scaler was fitted only on the training data and subsequently applied to the validation and test data.

The time series was then converted into a supervised learning structure using a 5-day sliding window, in which the five preceding observations were used to predict precipitation on the following day. The treatment of sequences at the boundaries between training, validation, and test periods should be specified to clarify how historical observations were used when generating validation and test sequences.

Table 1 presents the descriptive statistics of the meteorological variables. Precipitation had a positive skewness of 3.28, indicating a strongly right-skewed distribution with a heavy tail. The substantially higher mean precipitation than median precipitation also indicates the presence of relatively large precipitation values among predominantly low-precipitation observations.

Table 1. Descriptive Statistics of Meteorological Variables

Variable	Mean	Median	Std. Dev.	Skewness
Air temperature (°C)	27.70	27.80	0.95	-0.46
Relative humidity (%)	78.40	79.10	6.12	-0.32
Air pressure (hPa)	1010.15	1010.20	1.72	0.08
Wind speed (m/s)	3.42	3.30	1.18	0.54
Precipitation (mm)	6.85	0.20	14.72	3.28

Data splitting was performed using a time-series approach to preserve the temporal structure of the observations and reduce the risk of temporal leakage. Unlike random splitting, which may place observations from different periods into the training and evaluation subsets, the time-series approach assigns earlier observations to training and later observations to validation and testing. The split configurations are presented in Table 2.

Table 2. Time-Series Split Scheme

Split	Training Data	Validation Data	Test Data
Split 1	2019–2022	2023	2024
Split 2	2019–2023	2024	2025

3.3 eep Learning Model Architecture

The deep learning architectures evaluated in this study were selected to represent different approaches to modeling nonlinear relationships and temporal dependencies in tropical precipitation data. The three architectures were Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), and Temporal Convolutional Network (TCN).

1) Multilayer Perceptron (MLP)

The MLP model is a feedforward neural network consisting of three hidden layers with 64, 32, and 16 neurons, respectively. A Rectified Linear Unit (ReLU) activation function is applied to the hidden layers to model nonlinear relationships among the meteorological variables. Batch normalization and dropout with rates of 0.3 and 0.2, respectively, are incorporated to support training stability and reduce overfitting. The MLP operation can be expressed as:

$$y = f(W_2 \cdot g(W_1 \cdot x + b_1) + b_2) \quad (1)$$

Where x is the input vector, $W_1$and $W_2$are weight matrices, $b_1$and $b_2$are bias vectors, $g(\cdot)$ denotes the activation function in the hidden layer, and $f(\cdot)$ represents the output activation function (Aggarwal, 2018).

2) Long Short-Term Memory (LSTM)

The LSTM model uses a bidirectional architecture with a main layer containing 64 memory units. The bidirectional structure processes the input sequence in both directions within the observed input window. The LSTM layer is followed by batch normalization, dropout (0.3), a dense layer with 32 ReLU units, and dropout (0.2) before the output layer. The LSTM gating mechanism is formulated as:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4}$$

$$h_t = o_t \odot tanh\,(c_t) \tag{5}$$

Where f_t, i_t, and o_t denote the forget gate, input gate, and output gate, respectively, and C_t represents the cell state at time t.

3) Temporal Convolutional Network (TCN)

The TCN model consists of one-dimensional convolutional (Conv1D) blocks with 96 filters and a kernel size of 5. Dilated convolutions are applied using increasing dilation rates of 1, 2, 4, 8, 16, 32, and 64 to capture temporal dependencies over multiple time scales. Residual connections, spatial dropout (0.25), and batch normalization are incorporated into the architecture to support stable training. The causal convolution operation is defined as:

$$y_t = \sum_{i=0}^{k-1} f_i \cdot x_{t-d\cdot i} \tag{6}$$

Where f_i is the i-th convolution filter, k is the kernel size, and d is the dilation factor (Bai *et al.*, 2018).

3.4 Hyperparameter Tuning

Hyperparameter optimization was conducted using Grid Search and Random Search with time-based cross-validation following a rolling-window scheme. This approach was used to maintain the temporal structure of the data during model selection and reduce the risk of information leakage across periods.

Table 3. Hyperparameter Configuration for Each Model

Model	Parameter	Search Range	Tuning Method
MLP	Hidden layers, learning rate, optimizer	2–5; 0.001–0.01; Adam/RMSprop	Random Search
LSTM	Neurons, batch size, dropout, epochs	16–128; 16–64; 0.1–0.5; 50–200	Grid Search
TCN	Filters, kernel size, dilation, dropout	32–128; 2–5; 1–4; 0.1–0.3	Random Search

The Adam optimizer with a learning rate of 0.001 was selected based on preliminary experiments that indicated stable convergence. Model training used a batch size of 32 and an early-stopping mechanism based on validation loss to reduce overfitting and support generalization. The maximum number of epochs must be reconciled with the LSTM search range of 50–200 epochs in Table 3. The final training configuration and the tuning range should be stated consistently. The manuscript should also specify the number of search trials or candidate configurations, the validation objective used for model selection, and how the selected hyperparameters were determined for each time-series split.

3.5 Evaluation Metrics

Model performance was evaluated using four statistical metrics commonly used in time-series regression: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). RMSE assigns greater weight to larger prediction errors, MAE represents the average absolute prediction error, MAPE expresses prediction error as a percentage, and R^2 indicates the proportion of variance in the observed values explained by the model (Chai & Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|^2 \tag{8}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (10)$$

Where y_i denotes the actual value, $\hat{y}_i$ the predicted value, $\bar{y}$ the mean of the actual values, and n the number of observations. All computations were implemented in Python using the TensorFlow, Keras, NumPy, Pandas, and Scikit-learn libraries within a GPU-accelerated environment.

3.6 Evaluation and Validation Scheme

Model performance was measured using accuracy, precision, recall, and F1-score. ROC curve and AUC analysis was incorporated using the One-vs-Rest (OvR) approach for multi-class AUC calculation, with the Precision-Recall curve employed as a supplementary evaluation instrument — particularly relevant in security contexts where the cost of false negatives is asymmetric. A 5-Fold Cross-Validation scheme was applied to ensure generalization capability and freedom from overfitting, with sequential execution adopted to maintain memory efficiency without compromising validation integrity.

3.7 Tools and Technologies

The experimental implementation was developed in Python. Core libraries included Scikit-learn for Random Forest and TF-IDF vectorization, and the XGBoost library for the gradient boosting classifier. The prototype system was built using the Flask web framework, and feature mappings were serialized using pickle (char_vectorizer.pkl) to ensure full reproducibility across experimental runs.

4. Result and Discussion

4.1 Results

4.1.1 Climatological Characteristics of the Study Area

Monthly climatological analysis shows a distinct seasonal pattern at Soekarno–Hatta Airport, consistent with the tropical monsoon climate of western Java.

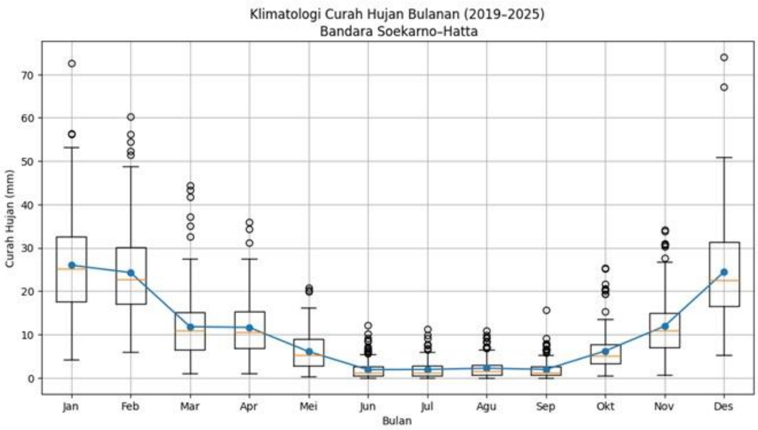

Figure 2. Boxplot of Monthly Rainfall Variability at Soekarno–Hatta Airport

Figure 2 shows that precipitation is generally highest from December to February, corresponding to the wet-season influence of the western monsoon in Java. During this period, the daily precipitation distribution exhibits greater variability and more extreme values, indicating a higher occurrence of heavy precipitation events. In contrast, the June–August period is characterized by lower precipitation and a narrower distribution, consistent with the dry-season conditions. Transitional periods, particularly March–May and October–November, show intermediate precipitation levels with considerable variability.

4.1.2 Validation of the Time-Series Split Approach

A comparison between random-split and time-series-split validation schemes is presented in Table 4. Random splitting produces a lower RMSE and higher R^2 than the time-series split. However, random splitting does not reproduce the chronological structure of operational forecasting because observations from different time periods may be distributed across the training and evaluation subsets. The time-series split instead preserves the temporal order of the observations and provides a more realistic assessment of out-of-sample performance for meteorological time-series prediction.

Table 4. Comparison of Validation Schemes for the TCN Model

Validation Method	RMSE (mm)	R^2
Random Split	8.95	0.91
Time-Series Split	10.32	0.87

The higher error and lower R^2 obtained with the time-series split indicate a more challenging but more realistic evaluation setting. The decrease in R^2 from 0.91 to 0.87 should therefore be interpreted in relation to the different validation designs rather than as evidence that the TCN model itself became less capable. The result supports the use of chronological validation when evaluating forecasting models on non-stationary meteorological time series.

4.1.3 Modeling Results and Performance Comparison

The training and validation loss curves show differences in the convergence behavior of the three models. As shown in Figure 3, the MLP training loss decreases from 0.66 to approximately 0.54, while the validation loss remains relatively stable at around 0.57–0.58. This indicates a limited improvement in validation performance during training. The LSTM model shows a simultaneous decline in both training and validation loss to approximately 0.54, indicating more consistent training and validation behavior. The TCN model shows the largest reduction in both losses, with training loss decreasing from 0.64 to 0.50 and validation loss decreasing from 0.58 to 0.51. The relatively small difference between the final training and validation losses indicates consistent convergence.

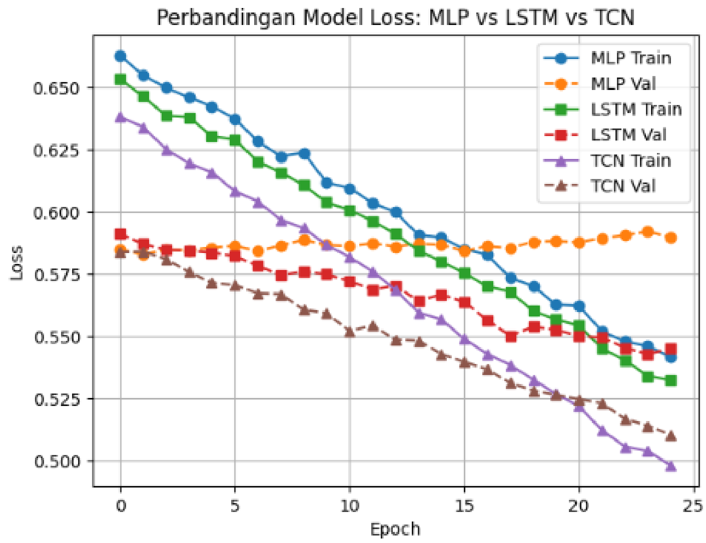

Figure 3. Comparison of Loss Curves for MLP, LSTM, and TCN Models

Table 5 summarizes the quantitative performance of the three models on the test dataset. The TCN model achieves the best overall prediction performance, with an RMSE of 10.32 mm, MAE of 7.15 mm, MAPE of 18.27%, and R^2 of 0.87. The LSTM model ranks second, with an RMSE of 12.45 mm, MAE of 8.72 mm, MAPE of 21.35%, and R^2 of 0.82. The MLP model produces the highest prediction errors, with an RMSE of 15.87 mm, MAE of 10.95 mm, MAPE of 28.64%, and R^2 of 0.74.

Table 5. Comparison of Precipitation Prediction Model Performance

Model	RMSE (mm)	MAE (mm)	MAPE (%)	R^2
MLP	15.87	10.95	28.64	0.74
LSTM	12.45	8.72	21.35	0.82
TCN	10.32	7.15	18.27	0.87

Figure 4 provides an additional visual comparison of the model predictions. The TCN predictions show the closest visual agreement with the observed precipitation series under typical conditions and during higher-precipitation periods. The LSTM predictions appear smoother than the observations, indicating that the model captures the general temporal pattern but does not reproduce some short-term fluctuations. The MLP predictions exhibit larger and less consistent deviations from the observed series, which is consistent with the absence of recurrent or temporal convolutional mechanisms for explicitly modeling temporal dependencies. The superior performance of the TCN may be associated with its use of causal and dilated convolutions, which allow the model to represent temporal dependencies across multiple time scales while preserving the chronological structure of the input sequence (Bai *et al.*, 2018). This result is also consistent with the broader use of convolutional architectures for temporal prediction reported in previous studies. The citation to Tao *et al.* (2019) should be verified before being retained as direct evidence for TCN performance.

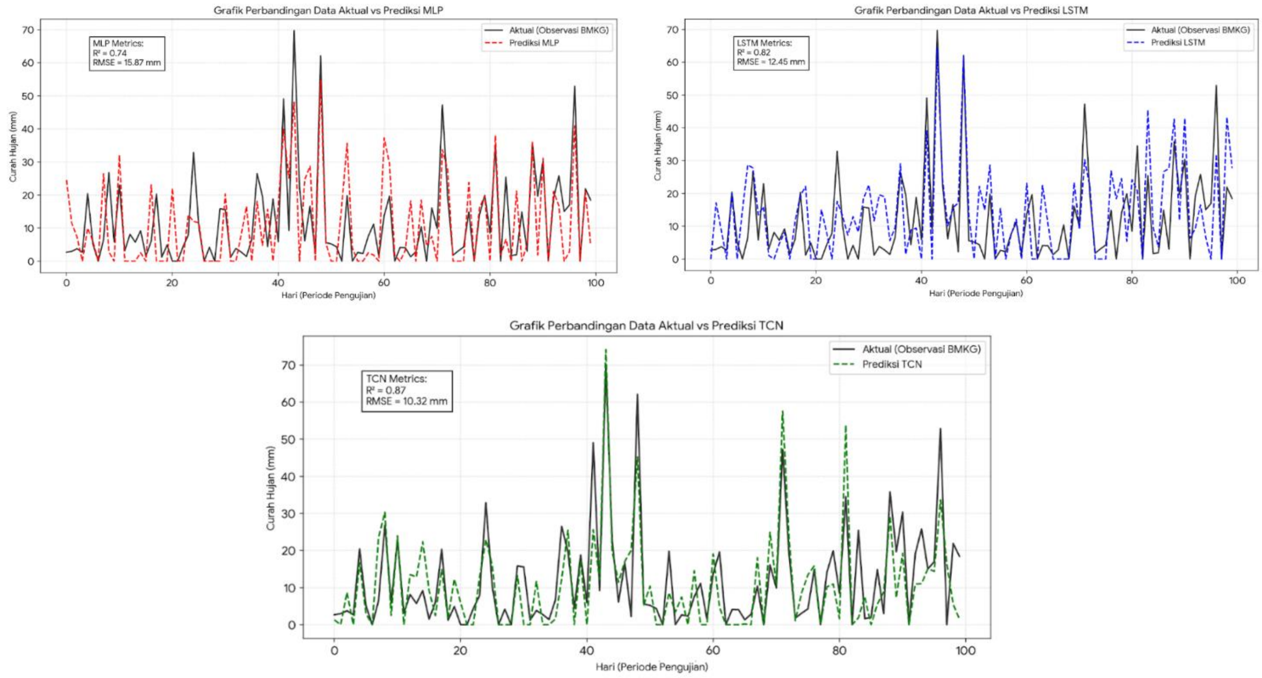

Figure 4. Prediction Results of MLP, LSTM, and TCN Models Compared with Actual Data

4.1.4 Sensitivity and Correlation Analysis

Permutation importance was used to examine the relative contribution of each meteorological input variable to precipitation prediction. The method measures the change in model performance after the values of a particular variable are randomly permuted. A larger degradation in model performance indicates a greater dependence of the model on that variable.

Table 6. Results of Input Variable Sensitivity Analysis

Variable	Contribution (%)	Influence Level
Humidity	32.5	Very High
Temperature	27.5	High
Air Pressure	18.4	Moderate
Wind Speed	12.6	Low
Other	8.7	Minor

Table 6 shows that relative humidity has the highest permutation importance, with a contribution of 32.5%, followed by air temperature at 27.5%. This indicates that the prediction model relies more strongly on information from humidity and temperature than from the other input variables. The strong contribution of humidity is physically plausible because atmospheric moisture availability is closely associated with condensation and precipitation processes (Guo *et al.*, 2024; Jun & Rind, 2024). Air temperature has the second-highest contribution, which may reflect its association with convective conditions and atmospheric instability. However, the sensitivity analysis does not establish a causal relationship between temperature and precipitation. Air pressure and wind speed show lower contributions in the model, although both variables may still provide information related to atmospheric circulation and moisture transport. The relatively low contribution of wind speed also indicates that its direct contribution to the model prediction is smaller than that of humidity and temperature under the conditions represented in the dataset.

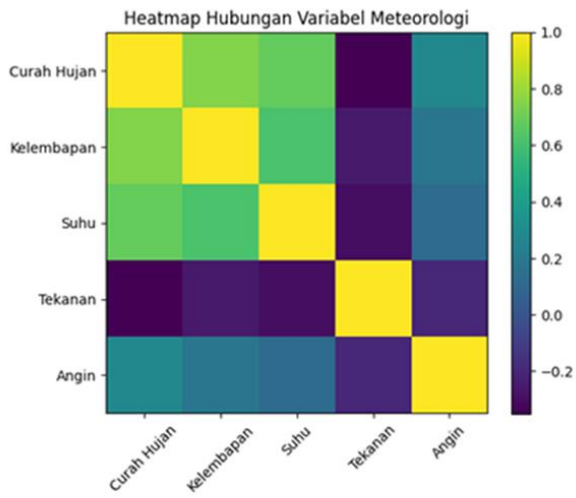

Figure 5. Heatmap of Correlations Between Meteorological Variables

Table 7. Correlation Matrix Between Meteorological Variables

Variable	Precipitation	Humidity	Temperature	Air Pressure	Wind
Precipitation	1.00	0.75	0.65	-0.30	0.20
Humidity	0.75	1.00	0.60	-0.20	0.10
Temperature	0.65	0.60	1.00	-0.25	0.05
Air Pressure	-0.30	-0.20	-0.25	1.00	-0.15
Wind	0.20	0.10	0.05	-0.15	1.00

Table 7 provides complementary information to the permutation-based sensitivity analysis. Precipitation shows a relatively strong positive correlation with humidity ($r = 0.75$) and temperature ($r = 0.65$), while its correlation with air pressure is negative ($r = -0.30$). The correlation between wind speed and precipitation is relatively weak ($r = 0.20$), indicating a limited linear association between these two variables in the observed dataset. The positive association between precipitation and humidity is consistent with the role of atmospheric moisture in precipitation formation. The negative association between pressure and precipitation may reflect the relationship between lower surface pressure and atmospheric conditions favorable to upward motion. However, these correlations represent statistical associations and should not be interpreted as evidence of causality. Similarly, the correlation results do not directly establish the same variable importance measured by permutation analysis because the two methods evaluate different properties of the data and model.

4.1.5 Model Performance Under Seasonal Climate Variability

To assess model performance under different seasonal conditions, the evaluation was conducted separately for the dry and wet seasons.

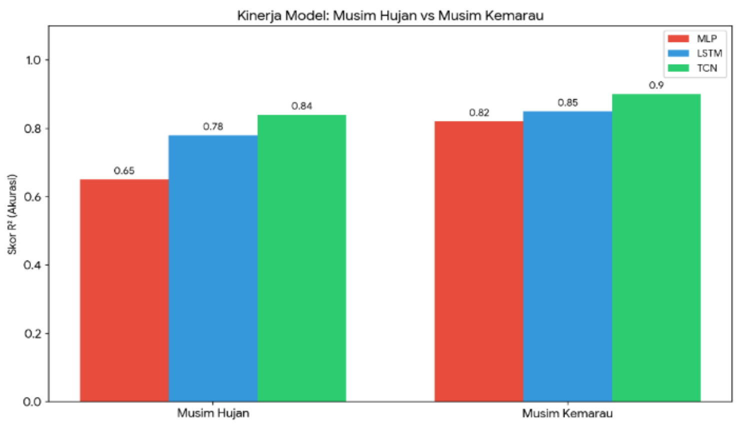

Figure 6. Comparison of Model Performance in Wet and Dry Seasons

Based on R^2, all three models perform better during the dry season than during the wet season. During the dry season, the R^2 values are 0.90 for TCN, 0.85 for LSTM, and 0.82 for MLP. During the wet season, these values decrease to 0.84, 0.78, and 0.65, respectively. The MLP model shows the largest reduction in R^2 between the two seasons, whereas the TCN model maintains the highest R^2 in both periods. These results indicate that TCN provides more consistent predictive performance across the seasonal conditions evaluated in this study. The definition of the wet and dry seasons and the number of observations in each period should be reported in the Methodology or in the figure/table description.

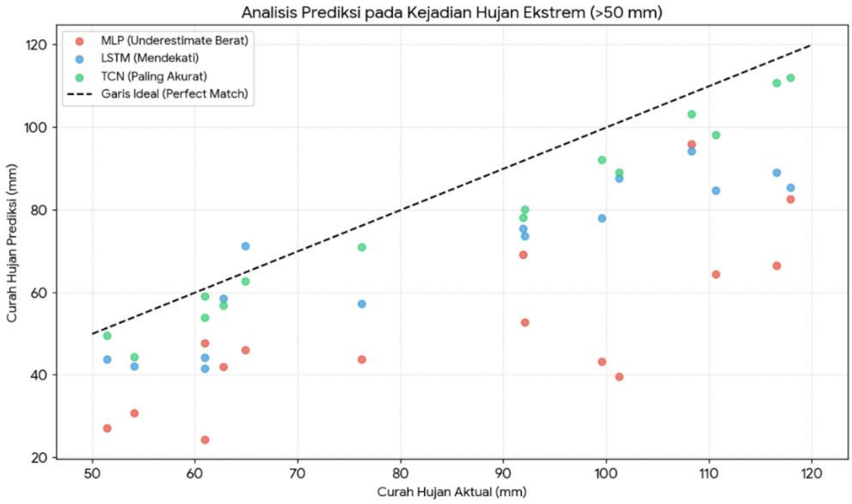

Figure 7. Model Performance for Extreme Precipitation Events

Figure 7 provides a visual comparison of the model predictions for extreme precipitation events. The TCN predictions appear closest to the 1:1 reference line, while the LSTM predictions show a tendency to underestimate the highest precipitation values. The MLP model exhibits larger deviations from the reference line, particularly at higher precipitation levels. These patterns indicate that TCN is better able to reproduce high-precipitation observations within the extreme-event subset evaluated in this study. The lower prediction accuracy during extreme precipitation events may be associated with the nonlinear and highly variable nature of tropical precipitation. Heavy precipitation can result from interactions among local convection, atmospheric moisture, circulation, and larger-scale climate variability. Such conditions can be difficult for statistical learning models to represent, particularly when extreme observations are relatively infrequent in the training data (Das *et al.*, 2022; Jiang *et al.*, 2024).

4.2 Discussion

The results demonstrate that the Temporal Convolutional Network (TCN) provides the best overall performance for daily precipitation prediction at Soekarno–Hatta International Airport, followed by LSTM and MLP. The TCN achieved an RMSE of 10.32 mm, MAE of 7.15 mm, MAPE of 18.27%, and R^2 of 0.87, indicating that the model was able to explain a substantial proportion of the variability in daily precipitation. The LSTM also showed relatively strong performance, whereas the MLP produced the largest prediction errors. The differences among the three architectures suggest that explicitly modeling temporal information is important for precipitation forecasting in the study area. The stronger performance of TCN can be associated with its causal and dilated convolutional structure. Unlike MLP, which processes the input through fully connected layers without an explicit temporal mechanism, TCN can examine patterns across different temporal scales through dilated convolutions. The use of increasing dilation rates allows the model to access observations from increasingly wider temporal intervals while maintaining the chronological order of the input sequence. This characteristic is particularly relevant to precipitation because rainfall conditions on a given day may be associated with atmospheric conditions observed during previous days. The result is consistent with the general capability of TCN architectures to model long-range temporal dependencies in sequential data (Bai *et al.*, 2018). The specific studies cited elsewhere in the manuscript as evidence for TCN superiority should be verified before being used as direct comparisons.

The LSTM also demonstrated good predictive capability, with an R^2 of 0.82. Its performance indicates that recurrent modeling can capture temporal dependencies in the meteorological observations. However, the lower performance compared with TCN suggests that the recurrent structure used in this study may not represent the temporal patterns as effectively as the dilated convolutional structure. The smoother LSTM predictions observed in the prediction plots may also indicate that the model captures the general precipitation pattern while failing to reproduce some short-term fluctuations and high-intensity events. This behavior is relevant because precipitation is characterized by substantial variability and a relatively small number of high-intensity observations.

The MLP produced the lowest performance among the three models, with an R^2 of 0.74 and an RMSE of 15.87 mm. The result does not indicate that MLP is unsuitable for meteorological prediction in general. Rather, under the experimental configuration used in this study, the feedforward architecture appears less effective than the sequential architectures for representing temporal precipitation patterns. The five-day sliding window provides historical observations to the MLP, but the architecture does not explicitly model temporal dependencies through recurrent or temporal convolutional mechanisms. This limitation may contribute to the larger deviations observed between the MLP predictions and the actual precipitation values.

The validation results further emphasize the importance of preserving chronological order when evaluating precipitation forecasting models. The TCN produced an RMSE of 8.95 mm and an R^2 of 0.91 under random splitting, compared with an RMSE of 10.32 mm and an R^2 of 0.87 under the time-series split. The difference indicates that random splitting produces more favorable evaluation results than chronological validation. For operational forecasting, however, the model must predict future observations using information available from previous periods. Therefore, the time-series split provides a more realistic representation of the forecasting problem. The reduction in performance should not be interpreted simply as deterioration of the TCN model, but as evidence that chronological evaluation presents a more demanding test of its ability to generalize to later observations.

The climatological analysis also provides an important explanation for the differences in model performance between seasons. Precipitation at Soekarno–Hatta shows clear seasonal variability, with greater precipitation and wider distributions during the wet season and lower precipitation during the dry season. The R^2 values of all three models decrease during the wet season, with the largest reduction occurring in the MLP model. This result indicates that prediction becomes more difficult when precipitation variability increases. The TCN maintains the highest R^2 in both seasons, suggesting that its temporal convolutional structure provides relatively consistent performance under different seasonal conditions. However, this interpretation is based on the reported R^2 values, and further evaluation using RMSE and MAE for each season would provide a more complete assessment.

The sensitivity analysis provides additional information regarding the meteorological variables used by the models. Relative humidity has the highest permutation importance at 32.5%, followed by temperature at 27.5%. This result is physically plausible because atmospheric moisture availability is closely associated with precipitation formation. A higher level of atmospheric moisture provides conditions that are more favorable for condensation and cloud development. The importance of humidity in the prediction model is therefore consistent with the physical role of water vapor in precipitation processes. However, permutation importance should be interpreted as a measure of the model's dependence on an input variable rather than as evidence of a causal relationship.

Air temperature was the second most influential variable. Its contribution may be related to the association between temperature and atmospheric instability and convective activity. In tropical environments, surface heating can contribute to the development of convective conditions, particularly when sufficient atmospheric moisture is available. Nevertheless, the positive correlation between temperature and precipitation reported in this study should not be interpreted as direct evidence that higher temperature causes higher precipitation. Other atmospheric processes may influence this relationship.

Air pressure and wind speed had lower permutation importance than humidity and temperature. The relatively weak correlation between wind speed and precipitation also indicates that surface wind speed has a limited linear association with precipitation in the observed dataset. However, this does not mean that wind-related processes are unimportant for precipitation formation. Wind direction, moisture transport, convergence, and vertical atmospheric motion are not directly represented by the single wind-speed variable used in this study. Therefore, the relatively low importance of wind speed may partly reflect the limited representation of atmospheric circulation in the input variables. If wind direction or other circulation variables are available, their inclusion could be examined in future studies.

The correlation analysis complements the permutation-based sensitivity analysis but should not be interpreted as a direct validation of variable importance. The relatively strong positive correlations between precipitation and humidity ($r = 0.75$) and between precipitation and temperature ($r = 0.65$) indicate that these variables have substantial linear associations with precipitation in the dataset. In contrast, the correlation between precipitation and wind speed is relatively weak ($r = 0.20$). The difference between correlation and permutation importance is important because correlation evaluates pairwise statistical association, whereas permutation importance evaluates the effect of disrupting a variable on model prediction performance. Therefore, the two analyses provide different but complementary perspectives on the meteorological predictors.

The analysis of extreme precipitation events indicates that TCN provides the closest predictions to the 1:1 reference line, while LSTM tends to underestimate the highest precipitation values and MLP shows larger deviations. This result is important because high-intensity precipitation events are particularly relevant to airport operations. However, extreme precipitation is relatively less frequent than ordinary precipitation, which makes it more difficult for a regression model trained on the overall distribution to learn these cases accurately. The ability of TCN to produce predictions closer to the observed extreme values may be related to its ability to represent temporal patterns over multiple scales. Nevertheless, the manuscript should explicitly define the threshold used to classify an observation as an extreme precipitation event and report the number of observations included in this subset.

From an operational perspective, the findings indicate that TCN is the most promising architecture among the three models evaluated for daily precipitation prediction at Soekarno–Hatta Airport. Its higher performance across the overall test set and its relatively stable performance between wet and dry seasons suggest potential for use as a supporting component in airport weather forecasting. However, the model should not be interpreted as a replacement for operational meteorological forecasting systems. The current study uses a limited set of surface meteorological variables and daily observations, whereas airport weather conditions can also be influenced by variables such as wind direction, solar radiation, cloud characteristics, atmospheric instability, radar observations, and larger-scale climate indices. These additional variables could potentially improve the representation of precipitation processes, particularly for extreme events.

The findings support the use of temporal deep learning architectures for precipitation forecasting at tropical airport locations. TCN achieved the best predictive performance among MLP, LSTM, and TCN under the experimental conditions used in this study. The results also show that model evaluation should preserve the chronological structure of meteorological observations because random splitting can produce more optimistic performance estimates. Humidity and temperature were the variables to which the model was most sensitive, while seasonal and extreme-event analyses showed that precipitation prediction remains more challenging under highly variable conditions. Future research should therefore focus on expanding the input variables, improving the representation of extreme precipitation events, and evaluating the model using longer forecasting horizons and additional airport meteorological observations.

5. Conclusion and Recommendations

This study developed and compared three deep learning architectures, namely MLP, LSTM, and TCN, for daily precipitation prediction at Soekarno–Hatta International Airport. The time-series split provided a more realistic evaluation of model performance by preserving the chronological structure of the observations. Among the three architectures, TCN achieved the best performance, with an RMSE of 10.32 mm, MAE of 7.15 mm, MAPE of 18.27%, and R^2 of 0.87. LSTM ranked second, with an RMSE of 12.45 mm, MAE of 8.72 mm, MAPE of 21.35%, and R^2 of 0.82, while MLP obtained an RMSE of 15.87 mm, MAE of 10.95 mm, MAPE of 28.64%, and R^2 of 0.74. The sensitivity analysis indicated that relative humidity was the most influential input variable, followed by air temperature, while air pressure and wind speed showed lower permutation importance. These results are consistent with the role of atmospheric moisture and thermal conditions in precipitation processes. However, the sensitivity values should be interpreted as the contribution of each variable to model predictions rather than as evidence of a direct causal relationship with precipitation. The reported contribution of air temperature should also be verified because Table 6 reports 27.5%, whereas the original conclusion reports 27.8%. The seasonal evaluation showed that all three models performed better during the dry season than during the wet season. Model performance also decreased when predicting extreme precipitation events, indicating that the heavy-tailed and highly variable characteristics of tropical precipitation remain challenging for the evaluated architectures. TCN maintained the strongest performance across the evaluated conditions, suggesting that its temporal convolutional structure was more effective than the MLP and LSTM configurations used in this study.

Based on these findings, TCN can be considered the most effective architecture among the three models evaluated for daily precipitation prediction at Soekarno–Hatta International Airport. However, the results do not yet establish its effectiveness in an operational forecasting environment because real-time operational testing was not conducted. Therefore, its application to airport forecasting and early-warning systems should be considered a potential implementation rather than a demonstrated operational capability. Future research may investigate hybrid architectures that combine TCN with attention mechanisms or Transformer-based models to improve the representation of complex temporal patterns, particularly during extreme precipitation events. Additional predictors, such as ENSO indices and solar radiation, may also be incorporated to represent broader atmospheric and climatic influences. Further evaluation using longer forecasting horizons, additional meteorological variables, and real-time operational testing would provide a stronger basis for assessing the practical applicability of the proposed approach.

References

Aggarwal, C. C. (2018). *Neural networks and deep learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0

Akhtar, R. (2020). Conclusion and suggestions. In *Extreme weather events and human health* (pp. 371–376). Springer International Publishing. https://doi.org/10.1007/978-3-030-23773-8_25

Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*. arXiv. https://arxiv.org/abs/1803.01271

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chamatidis, I., Tzanes, G., Istrati, D., Lagaros, N. D., & Stamou, A. (2023). Short-term forecasting of rainfall using sequentially deep LSTM networks: A case study on a semi-arid region. *Environmental Sciences Proceedings, 26*(1), 157. https://doi.org/10.3390/environsciproc2023026157

Chang, C.-Y., Lu, C.-W., & Wang, C.-J. (2021). A multi-step-ahead Markov conditional forward model with cube perturbations for extreme weather forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(8), 6948–6955. https://doi.org/10.1609/aaai.v35i8.16856

Das, S., Kamruzzaman, M., & Islam, A. R. Md. T. (2022). Assessment of characteristic changes of regional estimation of extreme rainfall under climate change: A case study in a tropical monsoon region with the climate projections from CMIP6 model. *Journal of Hydrology, 610*, 128002. https://doi.org/10.1016/j.jhydrol.2022.128002

de Jesús Machado-Guillén, G., Cruz-Duarte, J. M., Conant-Pablos, S. E., & Filus, K. (2024). Harnessing machine learning for reliable weather forecasting: Meteorological impact on sustainable energy in Monterrey. *2024 IEEE Congress on Evolutionary Computation (CEC)*, 1–8. https://doi.org/10.1109/CEC60901.2024.10611993

Ding, Y. H., Liu, Y. J., Xu, Y., & Chen, Y. (2024). The impact of climate warming on weather forecasting and coping strategies. *Climate Change Research, 20*(6), 682–688.

Guo, Q., Chen, Y., Miao, X., & Hao, Y. (2024). The response of cloud precipitation efficiency to warming in a rainfall corridor simulated by WRF. *Atmosphere, 15*(11), 1381. https://doi.org/10.3390/atmos15111381

Hudnurkar, S., Wanchoo, A., & Malhotra, A. (2015). Optimized artificial neural network model for day ahead maximum temperature forecasting. *International Journal of Applied Engineering Research, 10*(19), 40647–40655.

Jeong, C.-H., & Yi, M. Y. (2023). Correcting rainfall forecasts of a numerical weather prediction model using generative adversarial networks. *The Journal of Supercomputing, 79*(2), 1289–1317. https://doi.org/10.1007/s11227-022-04686-y

Jiang, Q., Cioffi, F., Li, W., Tan, J., Pan, X., & Li, X. (2024). Hybrid multilayer perceptron and convolutional neural network model to predict extreme regional precipitation dominated by the large-scale atmospheric circulation. *Atmospheric Research, 304*, 107362. https://doi.org/10.1016/j.atmosres.2024.107362

Jun, T., & Rind, D. (2024). Increasing contribution of the atmospheric vertical motion to precipitation in a warming climate. *Communications Earth & Environment, 5*(1), 541. https://doi.org/10.1038/s43247-024-01676-1

Lee, K., Kang, J., Paing, Z. M., & Baik, H. (2023). Prediction of aircraft delay at busy airport considering weather information with machine learning techniques. *2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC)*, 1–6. https://doi.org/10.1109/DASC58513.2023.10311121

Ma, H., Wang, J., Han, B., Zhang, Z., & Bao, H. (2024). Temporal convolutional network with soft threshold and contractile self-attention mechanism for remaining useful life prediction of rolling bearings. *Measurement Science and Technology, 35*(12), 126004. https://doi.org/10.1088/1361-6501/ad73f1

Maheng, D., Bhattacharya, B., Zevenbergen, C., & Pathirana, A. (2024). Changing urban temperature and rainfall patterns in Jakarta: A comprehensive historical analysis. *Sustainability, 16*(1), 350. https://doi.org/10.3390/su16010350

Meheretie, D. L., Ji, B., Mogadem, M. M., & Tamir, T. S. (2021). Modeling of neural networks for weather forecasting in Debre Markos, Ethiopia. *Journal of the Chinese Institute of Engineers, 44*(8), 762–769. https://doi.org/10.1080/02533839.2021.1977715

Nikam, S. K., & Mane, S. B. (2022). Weather forecasting using long short-term memory. In A. Kumar & S. Mozar (Eds.), *Proceedings of the International Conference on Computing, Communication, Control and Automation (ICCCE 2021)* (pp. 681–687). Springer. https://doi.org/10.1007/978-981-16-7985-8_71

Nugraha, Y. E. N., Ariawan, I., & Arifin, W. A. (2023). Weather forecast from time series data using LSTM algorithm. *Jurnal Teknologi Informasi dan Komunikasi, 14*(1), 144–152. https://doi.org/10.51903/jtikp.v14i1.531

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems, 12*(11). https://doi.org/10.1029/2020MS002203

Tao, Q., Liu, F., Li, Y., & Sidorov, D. (2019). Air pollution forecasting using a deep learning model based on 1D ConvNets and bidirectional GRU. *IEEE Access, 7*, 76690–76698. https://doi.org/10.1109/ACCESS.2019.2921578

Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems, 12*(9). https://doi.org/10.1029/2020MS002109