Vol. 6 No. 2 (2025), Page 263-272

GEOGRAPHIA
Jurnal Pendidikan dan Penelitian Geografi

ISSN: 2774-6968

DAILY RAINFALL FORECAST BASED ON MULTI-STATION OBSERVATION DATA IN MEDAN CITY

Nora Valencia Sinaga[1,2], Aulia Siti Aisjah[1*], Syamsul Arifin[1], Immanuel Jhonson Arizona Saragih[2]

[1]Department of Physics Engineering, Faculty of Industrial Technology and Systems Engineering, Sepuluh Nopember Institute of Technology, Surabaya 44101, Indonesia
[2]Department of Meteorology, Indonesia Agency for Meteorology, Climatology, and Geophysics (BMKG), Medan 23011, Indonesia

Email: nora.sinaga@bmkg.go.id[1], auliasa@ep.its.ac.id[2*]

Website Jurnal: http://ejurnal.unima.ac.id/index.php/geographia

DOI: https://doi.org/10.53682/82y51t67
(Accepted: 19-11-2025; Revised: 28-12-2025; Approved: 14-01-2026)

ABSTRACT
This study develops a multi-horizon daily rainfall forecasting model using the Long Short-Term Memory (LSTM) deep learning method, based on multi-station Automatic Weather Station (AWS) data in Medan City. Ten-minute AWS data from multiple stations (2021–2024) were merged and time-synchronized (UTC), followed by a quality control process including physical range checks, rate-of-change filtering, inter-variable consistency checks, and spike detection. Missing values were addressed using linear interpolation for short gaps and Multiple Imputation by Chained Equations (MICE) for longer gaps. Predictor features were constructed from weather parameters (temperature, humidity, pressure, wind, radiation), aggregated to an hourly scale, and reshaped into input time windows for LSTM. A two-layer LSTM model (128–64 units, 0.3 dropout, Adam optimizer) was trained to predict daily rainfall up to five days ahead. Evaluation metrics, including RMSE, MAE, POD, FAR, and CSI (with rainfall threshold ≥1 mm/day), indicated strong model performance: for instance, RMSE was below 10 mm/day for 1–3 day horizons, with POD above 0.80 and FAR below 0.20. The LSTM model outperformed conventional statistical models, yielding an accuracy improvement of approximately 30–40%. These findings highlight the potential of high-resolution AWS-based automatic forecasting systems to support hydrometeorological disaster mitigation in tropical urban areas.

Keywords: *Automatic weather station, Daily rainfall, Deep learning, Long short-term memory, Medan city, Multi-horizon forecasting.*

INTRODUCTION

Rainfall is a component of the climate and hydrological system, playing a crucial role in balancing water resources, supporting agriculture, and managing hydrometeorological disaster risk. In tropical regions like Indonesia, rainfall patterns vary significantly spatially and temporally due to the influence of regional atmospheric such as the Asia-Australia monsoon and the Intertropical Convergence Zone (ITCZ), global phenomena such as El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) (Aldrian & Dwi Susanto, 2003). This variability makes rainfall prediction at daily and local scales, both a scientific challenge and a practical exercise.

Medan, as the largest economic center in North Sumatra, has a complex climatological character influenced by the topography of the east coast of Sumatra and by its massive urbanization. The urban heat island phenomenon and local circulation enhance the formation of convective clouds, which have the potential to produce intense rainfall (Hidayat & Soekirno, 2021; Lubis, Situmorang, et al., 2025). The major flooding event on November 18, 2022, caused by extreme rainfall and the overflowing of the Deli River, inundating nine sub-districts with water levels of 10-70 cm (Ministry of Health of the Republic of Indonesia, 2022), underscored the urgency of accurate rainfall forecasting for urban disaster mitigation.

Conventional statistical methods, such as ARIMA and linear regression, have limitations in capturing nonlinear relationships and long-term dependencies between meteorological parameters (IPCC), 2023). In contrast, deep learning approaches such as Long Short-Term Memory (LSTM) can learn complex temporal patterns and long-term memory from time series data (Gao et al., 2022; Hochreiter & Schmidhuber, 1997; Lubis, Ghazali, et al., 2025). LSTM has been shown to improve rainfall prediction accuracy compared to classical models (Barrera-Animas et al., 2022). Furthermore, the availability of high-resolution data from Automatic Weather Stations (AWS) opens the door to developing prediction models based on actual observations. AWS records meteorological parameters such as temperature, humidity, air pressure, wind speed and direction, and solar radiation at 10-minute intervals, thereby capturing micro-atmospheric dynamics relevant to short-term forecasting (Cheng et al., 2023; Muhajir et al., 2021; Shaw, 2024; Yuan, 2025).

However, the application of LSTM models trained on multi-station AWS data to predict daily rainfall in tropical urban areas of Indonesia remains rare. Most previous studies have used only single data sets and failed to account for spatial relationships among stations, even though spatial correlation can enrich the representation of local atmospheric systems. This study attempts to fill this gap by designing an LSTM model based on multi-station AWS data in Medan City to forecast daily rainfall up to several days in advance (multi-horizon). The 10-minute-resolution AWS data were combined, quality-controlled, imputed using Multiple Imputation by Chained Equations (MICE), and transformed into time windows (windowing) for model input. The prediction results are then validated against daily Automatic Rain Gauge (ARG) data in the UTC time zone to measure the model's accuracy relative to actual observations.

Using this approach, the research aims to develop a rainfall forecasting system based on high-resolution observational data and deep learning that is adaptive to local conditions. The research's primary contribution lies in integrating multi-station spatial dimensions into a multi-output LSTM architecture, which is expected to improve the temporal representation and accuracy of daily rainfall forecasts in tropical urban areas such as Medan. In addition to providing scientific contributions to the application of Deep Learning to hydrometeorology, this research is expected to support disaster early warning systems and operational water resource governance.

DATA AND METHODS

This research uses a quantitative experimental approach based on data-driven modeling, focusing on the development and evaluation of a Long Short-Term Memory (LSTM) Deep Learning predictive model for daily rainfall forecasting. Conceptually, the research design includes four main stages: data collection and preparation from a multi-station Automatic Weather Station (AWS) in Medan City; data preprocessing, including quality control (QC) and missing data imputation; sequential data generation for LSTM model training; and model performance evaluation against external test and validation data (Automatic Rain Gauge/ARG).

The overall process is summarized in Figure 1, which displays a flowchart of the research methodology from data acquisition to model validation. The research data are sourced from the AWS network operated by the Meteorology, Climatology, and Geophysics Agency (BMKG) in and around Medan. The data comprises four main stations: Stamet Kualanamu (STA5001), Staklim North Sumatra (STA2068), Stamar Belawan (STW1002), and AAWS Delser (STA3032). Each AWS records meteorological parameters at 10-minute intervals, including rainfall (rr, mm/10-minute), average air temperature (tt_air_avg, °C), relative humidity (rh_avg, %), air pressure (pp_air, hPa), wind speed (ws_avg, m/s), wind direction (wd_avg, °), and solar radiation (sr_avg, W/m²).

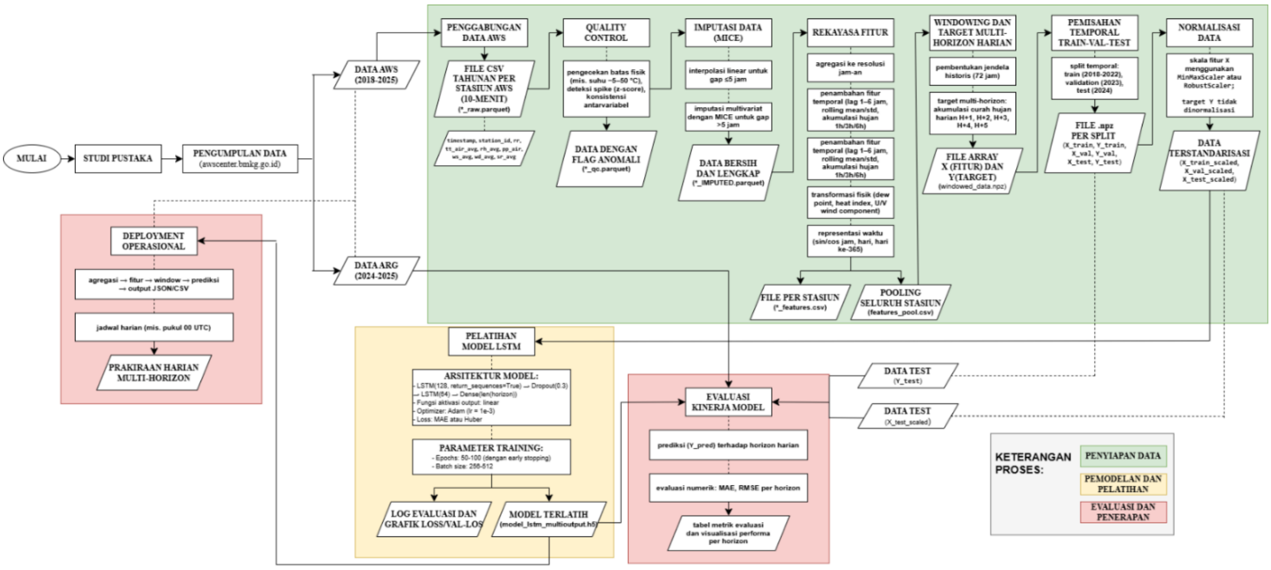

Figure 1. Research methodology flowchart: pre-processing stages, feature and sequence formation, LSTM model training, and evaluation and validation of results

Location metadata information for each station is presented in Table 1, which includes the latitude, longitude, and relative distance of each station to the center of Medan City. Figure 2 shows a map of the locations of the observation stations used in this study. To ensure data quality before use in model training, a multi-layered quality control (QC) process was performed following World Meteorological Organization (WMO) guidelines (Zahumenský, 2004). The QC stage includes range checks, spike detection, flatline tests to detect sensor interference, and consistency checks between variables (e.g., the difference between low humidity and high rainfall).

Tabel 1. Metadata station AWS in Medan city and its surroundings

ID	Station	Distance to Medan city (km)
STA5001	Stamet KNO	24.2 km
STA2068	Staklim Sumut	5.3 km
STW1002	Stamar Belawan	18.5 km
STA3032	AAWS Delser	30.2 km
150260	ARG Sunggal	9.0 km

Spike detection was implemented using a robust z-score method based on the median and Median Absolute Deviation (MAD), where values outside ±3MAD of the local median were identified as anomalies and replaced with missing values (NaN). This QC stage resulted in a 10-minute time series that was free of anomalies and ready for imputation.

After the data was declared valid, the missing data imputation process was performed to address observation gaps due to sensor or communication disruptions. Two approaches were applied sequentially: (1) time-based linear interpolation for short gaps (≤ 5 hours), and (2) Multiple Imputation by Chained Equations (MICE) for long gaps (> 5 hours). The MICE method was chosen because it maintains correlations between meteorological variables (Little & Rubin, 2019; Van Buuren & Groothuis-Oudshoorn, 2011).

The implementation was carried out using Iterative Imputer from scikit-learn with eight iterations and six closest predictor features. To ensure there was no statistical bias due to imputation, a Kernel Density Estimate (KDE) test was performed on the parameters of temperature, humidity, air pressure, and radiation using a distributional equality test. The results showed that the differences in distribution before and after imputation were relatively small, indicating that the data imputation process did not alter the statistical characteristics of the data.

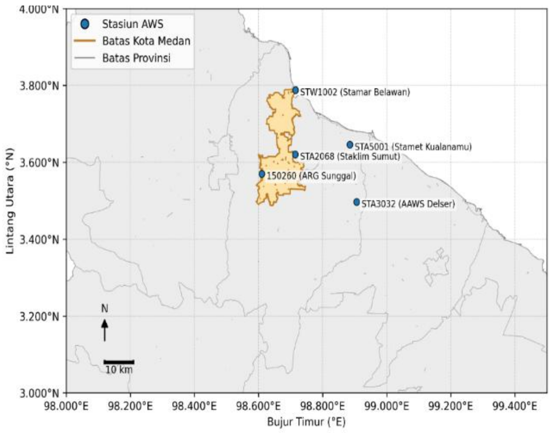

Figure 2. Map shows the research area.

The next step is the creation of a sequential dataset that serves as input for the LSTM model. The imputed data is converted into a windowed time series format, where each input sample consists of a sequence of meteorological data over the past 72 hours (equivalent to 432-times), and the output is the daily rainfall for the next five days (multi-horizon forecasting).

The choice of a 3-day input window is based on the consideration that tropical rainfall variability is generally influenced by atmospheric conditions up to 2–3 days in advance, while a 5-day horizon represents a statistically reliable short- to medium-term forecast period (Aldrian, 2008; Aldrian & Dwi Susanto, 2003; Estiningtyas et al., 2007; Yamanaka, 2016). The windowing is performed with a 3-hour stride to reduce redundancy. This process produces a three-part dataset: training (2021–2023), validation (late 2023), and test (2024).

Prior to training, all features were normalized using Robust Scaler, with the median and interquartile range (IQR) calculated from the training data only (Kuhn & Johnson, 2013). The Robust Scaler was chosen based on the skewed distribution of meteorological data and its high outlier count, making it more robust to extreme values than the Standard Scaler. The target output (daily rainfall) was not normalized to ensure the forecast results remained in physical units (mm/day) and easily interpreted.

A linear activation function was used in the output layer because the model's task was continuous regression. Optimization was performed using the Adam algorithm with an initial learning rate of 0.001 (Kingma & Cun, 2010), while the loss function used Mean Absolute Error (MAE). Training was conducted for 100 epochs with an early stopping mechanism (tolerance of 10 epochs) and Reduce LR On Plateau (a rate reduction of 0.5 if no improvement occurs within 4 epochs). A summary of the model structure and its parameters can be seen in Figure 3.

During training, the model received input from a pooled multi-station data dataset, assuming that weather conditions across stations in the Medan area have a high degree of spatial uniformity (Handoko et al., 1993). This allows the model to learn the city's general atmospheric patterns without losing significant local variations.

Model evaluation was conducted on the test data (2024) using MAE and Root Mean Square Error (RMSE) metrics to assess absolute and squared errors, as well as categorical metrics such as Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) with a rainfall threshold of ≥ 1 mm/day (Chattopadhyay & Chattopadhyay, 2014; Lubis, Simanjuntak, et al., 2025; Willmott & Matsuura, 2005). As a final step, external validation was conducted using daily rainfall data from the Automatic Rain Gauge (ARG) station of the Sunggal Water Company (PDAM) in Medan City to test the model's generalizability at locations not included in the training data.

RESULTS AND DISCUSSION

The results of this study cover data processing, model training, and performance evaluation of the Long Short-Term Memory (LSTM) model in predicting multi-horizon daily rainfall based on multi-station Automatic Weather Station (AWS) data in Medan City. The analysis process was conducted quantitatively on the model output, both on test data and external validation data.

Quality and Validity of Data

The initial phase of the research focused on ensuring the quality of the AWS data, which serves as the primary input for the LSTM model. Quality control results indicate that the AWS data used has an average completeness level above 98% after undergoing a review and imputation process. A summary of the completeness of the data for the 2021-2024 period is shown in Table 2, while a visualization of the comparison of completeness between stations for the 2018-2025 period is shown in Figure 4. STA1002 does not contain data from 2018-2020, so to optimize imputation results, only data from 2021-2024 was used. The results of the examination showed that from 2021 to 2024, most stations achieved completeness above 97%. Only one station (STW1002, Belawan) experienced significant data loss in 2021 (approximately 89% of data loss) due to sensor failure, but data returned to normal the following year. The pattern of data loss that emerged tended to be non-random and related to specific periods, indicating a type of missing not at random (MNAR). Therefore, the application of imputation based on Multiple Imputation by Chained Equations (MICE) was deemed appropriate, as it maintained relationships between meteorological variables without altering the statistical structure of the data (Little & Rubin, 2019).

Tabel 2. Recapitulation of AWS data completeness per station per year (2021–2024)

Stations	Year	Total Data (10-min)	% Data Valid	% Data Loss	Total Lines
STA2068	2021	41820	79.57	20.43	42870
STA2068	2022	51381	97.76	2.24	51398
STA2068	2023	51896	98.74	1.26	52302
STA2068	2024	51750	98.19	1.81	56510
STA3032	2021	31437	59.81	40.19	31458
STA3032	2022	52248	99.41	0.59	52248
STA3032	2023	52163	99.24	0.76	52163
STA3032	2024	51405	97.54	2.46	51405
STA5001	2021	50948	96.93	3.07	50948
STA5001	2022	52386	99.67	0.33	52386
STA5001	2023	52498	99.88	0.12	52498
STA5001	2024	52201	99.05	0.95	52201
STW1002	2021	5584	10.62	89.38	5584
STW1002	2022	49304	93.81	6.19	49304
STW1002	2023	52494	99.87	0.13	52494
STW1002	2024	52687	99.97	0.03	52687

In addition to data quantity, a missingness matrix was also examined. Figure 5 shows the temporal distribution of missing data (marked by dark gray blocks) throughout the period. The apparent pattern indicates that missingness is not purely random but rather clustered within specific periods. This indicates a type of non-random missingness (Missing Not at Random), necessitating gap filling to avoid bias toward specific seasons.

Evaluation of the imputation results shows that the daily rainfall distribution before and after imputation retains a right-skewed shape typical of tropical climates, with rainfall-free days predominating at more than 60% of the total. This is evident in Figure 6, which shows a histogram of daily rainfall before and after imputation. The similarity in the distribution shape indicates that the data imputation process did not shift the natural characteristics of the data, so the preprocessed dataset can be considered representative and reliable for training predictive models.

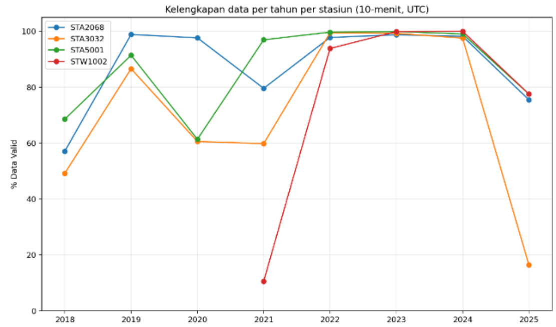

Figure 3. Percentage of AWS data completeness per year at each station (10-minute resolution, UTC time); significant improvements were seen in 2022–2024 at all stations compared to 2018–2020.

In addition to rainfall, the distribution of other meteorological variables (temperature, humidity, air pressure, and radiation) also remained consistent before and after the QC-imputation process. The Kernel Density Estimate (KDE) curves in Figure 5 show very small differences in key parameters, indicating that the data cleaning and imputation steps successfully maintained the overall statistical structure of the data. It can be seen that the "before" (blue) and "after" (orange) distribution curves overlap closely, with small differences in peaks or widths.

For example, the daily temperature distribution remains unimodal with a mean of around 27–28°C, the relative humidity distribution maintains a double-peaked shape (characteristic of morning and afternoon humidity patterns), and so on. Small improvements have occurred: the "after" imputation data tends to have a smoother curve and is more consistent across stations, after extreme anomalies are removed during the QC stage. This indicates that the applied QC and imputation methods are effective: they are able to fill in missing data without causing significant distortions to the statistical structure of the data. Thus, the pre-processed dataset is suitable for use as input for predictive models.

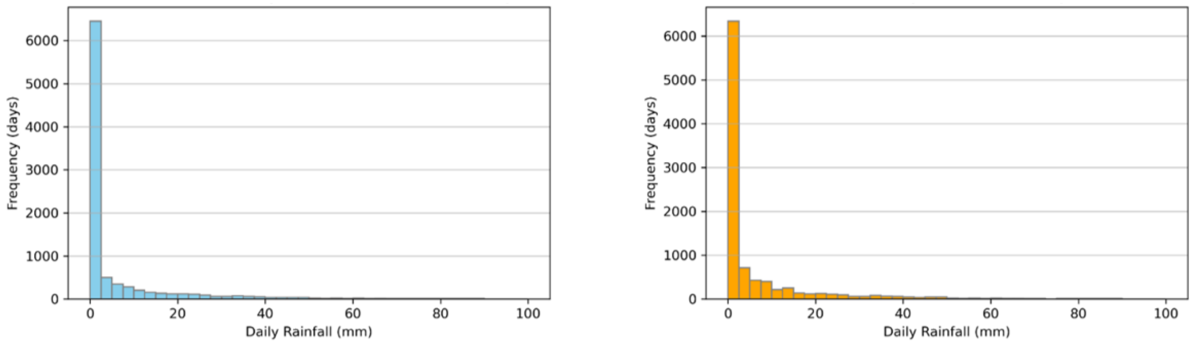

Figure 4. Histogram of daily rainfall distribution before (a) and after (b) the missing data imputation process

Sequential Data and Training

The pre-processed dataset is then converted into a sequential format for processing by the LSTM network. Each input sample consists of a sequence of meteorological data from the last 72 hours (432 time steps), while the target output is the total daily rainfall for the next five days. This scheme allows the model to learn the medium-term temporal relationship between past atmospheric conditions and future rainfall probability. A representation of this sequential data is shown in Figure 6, which shows the temporal variation of meteorological variables at one of the random observation stations.

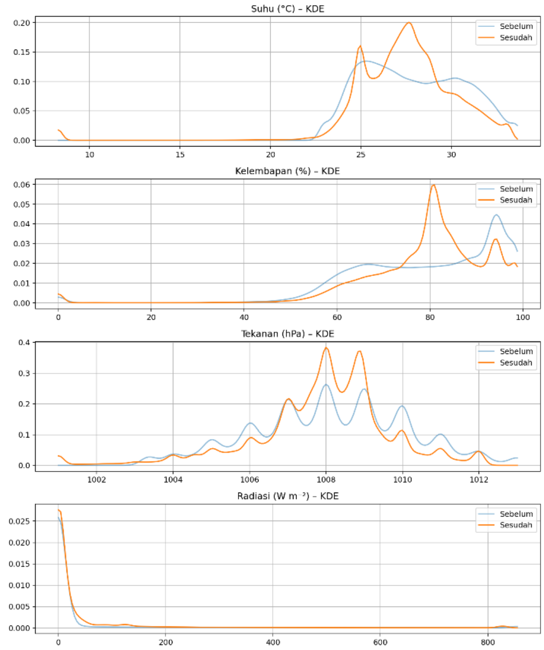

Figure 5. KDE graph of distribution of air temperature, humidity, pressure, and radiation, before vs after QC/imputation.

The model architecture consists of two consecutive LSTM layers (128 and 64 units), each with a dropout of 0.2 to prevent overfitting, and a Dense layer of five neurons representing the daily rainfall forecast for the next five days. The training process demonstrated stable convergence, with a Mean Absolute Error (MAE) of approximately 6.0 mm for the training data and approximately 8.3 mm per day for the validation data. The training and validation curves are shown in Figure 6, which show no significant differences between the two, indicating that the model does not suffer from overfitting and is capable of good generalization.

Overall, the results indicate that the LSTM model based on multi-station AWS data is capable of producing daily rainfall forecasts with moderate error and good interhorizon stability. The average MAE value of around 7–8 mm/day indicates that the model is sufficiently accurate for short-term operational applications, such as daily rainfall early warning systems. The consistency of the results up to a five-day horizon demonstrates that the LSTM's long-term memory is effective in recognizing temporal rainfall patterns in Medan. The model's performance in detecting extreme rainfall is still limited, largely due to the predominance of zero data (rainless days) and the relatively small number.

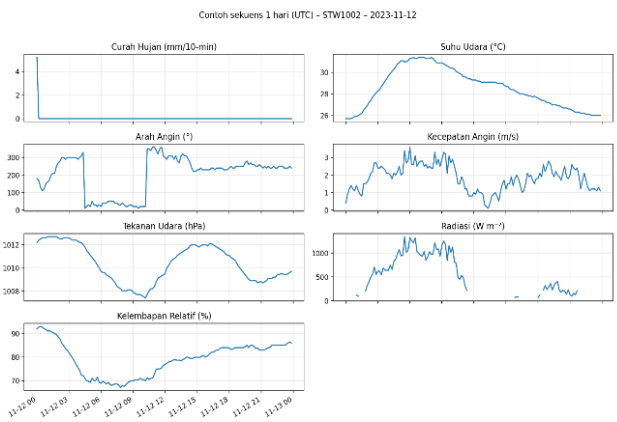

Figure 6. Example of daily meteorological data sequence representation (UTC) at station STW1002 on November 12, 2023.

These findings align with previous research showing that LSTM-based models outperform short-term quantitative predictions compared to extreme event predictions (Ghimire et al., 2021). Although external validation showed a performance decline, this opens up opportunities for further development. Integrating spatial variables (e.g., latitude–longitude coordinates or radar imagery) into the LSTM architecture has the potential to improve the model's ability to recognize spatial rainfall gradients (Gamboa-Villafruela et al., 2021).

Furthermore, the application of transfer learning across regions could expand the model's application to other cities with similar AWS networks. Theoretically, these research results strengthen the argument that deep learning approaches such as LSTMs can replace conventional statistical models (ARIMA, linear regression) in rainfall prediction in tropical regions with superior performance (Barrera-Animas et al., 2022). Practically, these research results demonstrate the potential for developing an automated rainfall forecasting system based on high-resolution observational data.

CONCLUSION

This research successfully developed a daily rainfall prediction model based on Long Short-Term Memory (LSTM) deep learning, utilizing high-resolution observation data from a multi-station Automatic Weather Station (AWS) in Medan City. The results demonstrate that the LSTM model is able to effectively learn temporal rainfall patterns and produce daily forecasts up to five days in advance with relatively low and stable error rates across horizons. The average Mean Absolute Error (MAE) of 7–8 mm/day and Root Mean Square Error (RMSE) of around 15 mm/day on the 2024 test data demonstrates that this approach has competitive predictive performance compared to conventional statistical methods such as ARIMA and linear regression. The application of a two-layer LSTM architecture with a 72-hour input window and multi-horizon output has proven effective in capturing temporal relationships between meteorological parameters, such as temperature, humidity, pressure, wind, and radiation. Furthermore, the Multiple Imputation by Chained Equations (MICE)-based quality control and imputation process applied to AWS data was proven to maintain the integrity of the data distribution without introducing significant statistical bias.

Although the model demonstrated good accuracy in quantitatively predicting light to moderate rainfall, the analysis also revealed a tendency to underestimate extreme rainfall events. External validation using Automatic Rain Gauge (ARG) data at the Sunggal Water Company (PDAM) showed a decrease in performance due to spatial differences between observation points and the training data, indicating the need for adjustments or spatial calibration if the model is applied to other locations. Thus, this study demonstrates that the use of LSTM based on multi-station AWS data can be a reliable approach for automated daily rainfall forecasting systems in tropical urban areas like Medan, while also providing an empirical basis for the development of hydrometeorological disaster mitigation systems based on high-resolution observational.

REFERENCES

(IPCC), I. P. on C. C. 2023. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781009157896

Aldrian, E. 2008. Meteorologi Laut Indonesia. *Jakarta: Badan Meteorologi Dan Geofisika*, 242–243.

Aldrian, E., & Dwi Susanto, R. 2003. Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *23*(12), 1435–1452.

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. 2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204.

Chattopadhyay, A. K., & Chattopadhyay, T. 2014. Time series analysis. In *Statistical Methods for Astronomical Data Analysis* (pp. 217–240). Springer.

Cheng, Z., Wang, L., & Yang, Y. 2023. A

hybrid feature pyramid CNN-LSTM model with seasonal inflection month correction for medium-and long-term power load forecasting. *Energies*, *16*(7), 3081.

Estiningtyas, W., Ramadhani, F., & Aldrian, E. 2007. Analisis Korelasi Curah Hujan Dan Suhu Permukaan Laut Wilayah Indonesia, Serta Implikasinya Untuk Prakiraan Curah Hujan (Studi Kasus Kabupaten Cilacap). *Agromet*, *21*(2), 46–60.

Gamboa-Villafruela, C. J., Fernández-Alvarez, J. C., Márquez-Mijares, M., Pérez-Alarcón, A., & Batista-Leyva, A. J. 2021. Convolutional lstm architecture for precipitation nowcasting using satellite data. *Environmental Sciences Proceedings*, *8*(1), 33.

Gao, W., Li, Z., Chen, Q., Jiang, W., & Feng, Y. 2022. Modelling and prediction of GNSS time series using GBDT, LSTM and SVM machine learning approaches. *Journal of Geodesy*, *96*(10).

Ghimire, S., Yaseen, Z. M., Farooque, A. A., Deo, R. C., Zhang, J., & Tao, X. 2021. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, *11*(1), 17497.

Handoko, I., Tania, J., & Rini, H. 1993. Klimatologi dasar: landasan pemahaman fisika atmosfer dan unsur-unsur iklim. *Jur. Geofisika Dan Meteorologi. FMIPA IPB. Bogor*.

Hidayat, D. J., & Soekirno, S. 2021. Development of temperature monitoring and prediction system for urban heat island (UHI) based on the internet of things. *Journal of Physics: Conference Series*, *1816*(1), 12054.

Hochreiter, S., & Schmidhuber, J. 1997. Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Kementerian Kesehatan RI 2022 Banjir di Kota-Medan, Sumatera-Utara, 18-11-2022, Pusat Krisis Kesehatan Kemkes RI. Tersedia pada: https://pusatkrisis.kemkes.go.id/Banjir-di-Kota-Medan-Sumatera-UtarA-18-11-2022-34 (Diakses: 22 Agustus 2025).

Kingma, D. P., & Cun, Y. 2010. Regularized estimation of image statistics by score matching. *Advances in Neural Information Processing Systems*, *23*.

Kuhn, M., & Johnson, K. 2013. *Applied predictive modeling* (Vol. 26). Springer.

Little, R. J. A., & Rubin, D. B. 2019. *Statistical analysis with missing data*. John Wiley & Sons.

Lubis, M. Z., Ghazali, M., Simanjuntak, A. V. H., Riama, N. F., Pasma, G. R., Priatna, A., Kausarian, H., Suryadarma, M. W., Pujiyati, S., & Simanungkalit, F. 2025. Decadal and seasonal oceanographic trends influenced by climate changes in the Gulf of Thailand. *The Egyptian Journal of Remote Sensing and Space Sciences*, *28*(2), 151–166.

Lubis, M. Z., Simanjuntak, A. V. H., Riama, N. F., Pasma, G. R., Dwinovantyo, A., Atmadipoera, A. S., Ansari, K., & Jamjareegulgarn, P. 2025. Numerical simulation of hydrodynamics and measurement in the Northwest Banda Sea. *Egyptian Journal of Aquatic Research*, *51*(1), 19–30.

Lubis, M. Z., Situmorang, E., Simanjuntak, A. V. H., Riama, N. F., Pasma, G. R., Dwinovantyo, A., Kausarian, H., Natih, N. M. N., Ansari, K., & Jamjareegulgarn, P. 2025. Indonesian Throughflow, spatial–temporal variability, and its relationship to ENSO events in the Lombok Strait. *Egyptian Journal of Aquatic Research*, *51*(1), 7–18.

Muhajir, M., Ismail, N., Syahreza, S., & Simanjuntak, A. V. H. 2021. Pemutakhiran Zona Musim (ZOM) Provinsi Aceh Menggunakan Data Blending Berbasis Non-Hirarki K-Means Clustering. *Jurnal Fisika Flux: Jurnal Ilmiah Fisika FMIPA Universitas Lambung Mangkurat*, *18*(1), 35–41.

Shaw, T. 2024. Automatic weather station. *2024 6th International Conference on Energy, Power and Environment (ICEPE)*, 1–5.

Van Buuren, S., & Groothuis-Oudshoorn, K. 2011. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Willmott, C. J., & Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82.

Yamanaka, M. D. 2016. Physical climatology of Indonesian maritime continent: An outline to comprehend observational studies. *Atmospheric Research*, *178*, 231–259.

Yuan, R. 2025. Rainfall prediction based on CNN-LSTM model under sliding window. *European Journal of Remote Sensing*, *58*(1), 2540106.

Zahumenský, I. 2004. Guidelines on quality control procedures for data from automatic weather stations. *World Meteorological Organization, Switzerland*, *955*, 2–6.