Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data Volume as the Primary Constraint in Tropical Rainfall Forecasting: A Comparative Analysis

Edward Alexandra[1], Manson Lie[2], Randy Liminson[3], Frans Mikael Sinaga[4]*
[1,2,3,4] Informatics Department, Faculty of AI and Data Sciences, Universitas Pelita Harapan, Indonesia
[1]03082230017@student.uph.edu, [2] 03082230003@student.uph.edu, [3] 03082230001@student.uph.edu, [4] frans.sinaga@uph.edu

Submitted :Mar 15, 2026 | **Accepted** : Apr 17, 2026 | **Published** : Apr 19, 2026

Abstract: North Sumatra's tropical climate requires accurate daily rainfall forecasting for paddy cultivation. This study evaluates three deep learning architectures — Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and CNN-BiLSTM — alongside XGBoost for daily rainfall forecasting in Deli Serdang Regency, North Sumatra, Indonesia. Two datasets were used: a two-year observational dataset from the Indonesian Climatological Station or known as BMKG (788 samples, January 2024 – March 2026) and an eleven-year NASA POWER reanalysis dataset (4,082 samples, January 2015 – March 2026), both at coordinates 3.6211°N, 98.7149°E. All models were evaluated using RMSE, MAE, MSE, and R^2 under a chronological 70/15/15% train-validation-test split. On BMKG dataset, all models achieved severely limited performance (highest R^2 = 0.0774, XGBoost), When trained on the larger NASA POWER dataset, all models exhibited consistent RMSE reductions of 63.89%-66.71%, with XGBoost achieving the best overall performance (RMSE = 7.5461 mm, R^2 = 0.1504) and the only positive R^2 across both datasets. Among deep learning architectures, BiLSTM consistently yielded the best R^2. The persistently low R^2 across all models reflects the fundamental challenge posed by zero-inflated, discontinuous nature of tropical rainfall distributions. A Wilcoxon signed-rank test confirmed that the performance differences between datasets were statistically significant for all models at $\alpha = 0.05$ (p = 0.0020–0.0371), with XGBoost and CNN-BiLSTM R^2 remaining significant after Bonferroni correction. These findings recommend BiLSTM as the promising deep learning candidate for future investigation with extended observational records.

Keywords: BiLSTM; daily rainfall forecasting; deep learning; Deli Serdang; NASA POWER; tropical climate; XGBoost

INTRODUCTION

Indonesia, as an agrarian nation, positions sector as a strategic national food commodity. It's geographic location between the Pacific and Indian Oceans exposes the country to monsoon systems and El Niño-Southern Oscillation (ENSO)-driven climate variability (Andrista et al., 2025), producing rainfall patterns that directly threaten agricultural productivity and making accurate climate prediction critical for planting schedule devisions (Ansari et al., 2023).

The consequences of rainfall prediction inaccuracies are particularly acute in North Sumatra, where most agricultural areas remain rain-fed. This vulnerability is evidenced by the ongoing need to optimize 225 hectares of dryland and the procurement of 931 water pump units over the past five years, rendering farmers highly vulnerable to climate anomalies, as further evidenced by the decline in harvest frequency resulting from flooding in 2025 (Direktorat Jenderal Prasarana dan Sarana Pertanian, 2025; Kurniawan, 2025). This issue is compounded by rainfall's inherently complex nature as a meteorological variable characterized by non-stationary and non-linear patterns. It is governed by intricate interactions between various atmospheric factors, making it challenging to model accurately using linear approaches. (Islam et al., 2024; Yuan, 2025).

Various efforts have been addressed rainfall-related challenges through advanced deep learning architectures (Jahangiri et al., 2025; Yuan, 2025). Tree-based models such as Random Forest and Decision Tree have demonstrated effectiveness in capturing complex non-linear patterns within climate-driven agricultural datasets

*name of corresponding author

(Islam et al., 2024), though they generally lack the capacity to explicitly model temporal dynamics. Consequently, Long Short-Term Memory (LSTM) was introduced to handle meteorological time series, proving more effective at retaining long-term dependencies (Yuan, 2025). Nevertheless, as a unidirectional model, LSTM processes sequences in a single direction from past to future, restricting its capacity to capture bidirectional temporal patterns within the data. Bidirectional LSTM (BiLSTM) addresses this limitation by processing sequences simultaneously in both forward and backward directions, achieving high accuracy in rainfall prediction tasks (Jiao & He, 2024), with recent studies reporting superior BiLSTM performance compared to conventional LSTM across various time series tasks (Jiang et al., 2023)

Despite these advances, three gaps remain unaddressed in existing literature. First, BiLSTM-based rainfall forecasting has been evaluated predominantly in non-tropical contexts — for instance, in lake basins of Shandong Province, China (Jiao & He, 2024) — leaving its effectiveness under Indonesia's tropical climate, which exhibits considerably greater complexity due to simultaneous interactions between ENSO, monsoon, and regional oceanographic factors (Audric Valennur et al., 2026; Ramadhan et al., 2024). Second, comparative evaluations of deep learning architectures against tree-based models such as XGBoost remain scarce in tropical and agricultural context, limiting our understanding of whether performance constraints arise from architectural inadequacy or data volume limitations. Third, no study has systematically evaluated these architectures using direct BMKG observational data from Deli Serdang Regency where precipitation forecasting accuracy has direct consequences for planting schedule optimization (Ansari et al., 2023).

Based on this identified gap, the present study proposes a BiLSTM-based rainfall prediction model directly evaluated on meteorological data from Deli Serdang Regency, North Sumatra, one of the strategic rice production centers contributing 10.8% of North Sumatra's total rice output (Badan Pusat Statistik, 2026). BiLSTM was evaluated alongside LSTM and CNN-BiLSTM to analyze comparative deep learning performance within Indonesia's tropical climate context. Furthermore, XGBoost was incorporated as a diagnostic model to empirically investigate whether performance limitations originate from architectural insufficiency or dataset volume constraints — an investigation further extended using an eleven-year NASA POWER dataset as conceptual evidence regarding minimum data requirements for accurate tropical rainfall modeling. Model performance was evaluated using RMSE, MAE, and R^2 metrics to produce a comprehensive assessment of prediction accuracy.

This study addresses these gap through three contributions: (1) it provides the first systematic empirical evaluation of BiLSTM, LSTM, CNN-BiLSTM, and XGBoost on BMKG observational data from Deli Serdang Regency, directly situating deep learning rainfall forecasting within Indonesia's tropical agricultural context; (2) it empirically isolates dataset volume by testing identical model configurations across two datasets of contrasting sizes — a two-year BMKG dataset and an eleven-year NASA POWER reanalysis dataset — under controlled experimental conditions; and (3) it characterizes the distributional challenge of zero-inflated tropical rainfall on regression-based models, providing a diagnostic foundation for future architecture design in this climate context.

LITERATURE REVIEW

Machine learning-based weather prediction has emerged as a rapidly growing field of study, driven by its capacity to handle high-dimensional datasets as well as to leverage historical and real-time data at a large scale. In general, the approaches found in the literature can be classified into two broad categories: traditional machine learning methods and deep learning methods (Zhang et al., 2025).

A study conducted in Bangladesh demonstrated that the Random Forest algorithm delivered the highest composite predictive performance among five evaluated models — including Gradient Boosting, Linear Regression, Decision Tree, and Neural Network — when forecasting Aman rice yields using meteorological inputs such as temperature and rainfall. These findings suggest that tree-based ensemble models exhibit strong robustness in handling non-linear relationships between climatic variables and agricultural outcomes (Islam et al., 2024). However, the fundamental limitation of traditional machine learning models lies in their inability to explicitly capture temporal dynamics, as conventional models tend to process each variable independently without accounting for inter-variable interactions within a temporal context. Furthermore, these models encounter difficulty in capturing more complex relationships within meteorological data, particularly when confronted with large scale inputs of high dimensionality (Zhang et al., 2025).

In response to these limitations, Recurrent Nural Network (RNN) architectures and subsequently Long Short-Term Memory (LSTM) were widely adopted for weather prediction tasks. LSTM was specifically designed to address the vanishing gradient problem inherent in standard RNN architectures, thereby enabling the retention of long-term information within temporal sequences (Jiang et al., 2023). Yuan (2025) demonstrated that LSTM-based models are effective for rainfall prediction, noting that LSTM networks are particularly suited for time series data due to their ability to capture long-term dependencies, and collectively highlighting the effectiveness of LSTM models in rainfall prediction through a review of prior work. The LSTM and BiLSTM architecture have transformed temporal modeling by integrating memory cells capable of capturing long-term dependencies within

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i2.16047 p-ISSN : 2541-044X

precipitation data (Jahangiri et al., 2025). However, rainfall is a phenomenon simultaneously influenced by various atmospheric indices, such that models that solely process sequences without integrating interactions among these variables will yield suboptimal predictions (Yuan, 2025). A further limitation of conventional LSTM lies in its unidirectional sequence processing, whereby the model is restricted to training signals in a single temporal direction and therefore cannot fully exploit the contextual information present in both past and future time steps within a sequence (Jiang et al., 2023).

Bidirectional LSTM (BiLSTM) emerged as a solution by processing temporal sequences simultaneously in two directions, namely forward and backward. (Jahangiri et al., 2025; Jiang et al., 2023). This approach enables the model to simultaneously integrate both past and future contextual information, thereby achieving context extraction from two temporal directions concurrently (Jiang et al., 2023). This mechanism operates through two parallel LSTM layers — a forward layer that processes sequences from past to future and a backward layer that processes sequences in the reverse direction — whereby the outputs of both layers are concatenated to produce a more comprehensive temporal representation (Jahangiri et al. 2025; Jiang et al. 2023). Several recent studies report that BiLSTM is capable of yielding strong performance in daily precipitation prediction, as evidenced by Jahangiri et al., which achieved an R^2 value of 0.638 for the BiLSTM component (Jahangiri et al., 2025), as well as Jiao & He, whose VMD-IPSO-BiLSTM hybrid model reported an NSE exceeding 0.88 in weekly rainfall prediction at the Nansi Lake Basin, China (Jiao & He, 2024).

Nevertheless, Indonesia's tropical climate is considerably more complex than the climate context of regions commonly studied in precipitation modeling, characterized by extreme rainfall fluctuations resulting from the interaction between ENSO variability and the Asian-Australian monsoon system (Audric Valennur et al., 2026). The El Niño phase generally triggers a reduction in rainfall across the majority of Indonesia's territory, while conversely, the La Niña phase induces extreme rainfall increases that have the potential to cause floods, landslides, droughts, and disruptions to clean water supply (Audric Valennur et al., 2026).This complexity is further compounded by other multivariate climate factors such as the shifting of the Intertropical Convergence Zone (ITCZ), Intra-seasonal variability, and regional oceanographic conditions that are inherently non-linear in nature (Audric Valennur et al., 2026). Consequently, the rainfall dynamics of tropical climate in Indonesia are fundamentally distinct from the semi-arid and temperate climate contexts commonly used in existing precipitation modeling studies, as reflected by the complexity of the ENSO interactions, monsoon dynamics, and local factors such as land-sea interactions that simultaneously influence precipitation patterns across Indonesia's maritime region (Audric Valennur et al., 2026; Ramadhan et al., 2024). The existing deep learning literature based on BiLSTM has thus far not afforded adequate attention to the complexity of this nature of tropical climate.

This condition is particularly relevant in regions such as Deli Serdang Regency, North Sumatra, which constitues one of the strategic rice production centres with a total production of 298,852 tons and a harvested area of 50,094 hectares in 2025 (Badan Pusat Statistik, 2026), However, rainfall anomalies may occur, as evidenced by BMKG's forecast for August 2025, which was categorized as 'Above Normal' with rainfall exceeding 115% of the 30-year historical average (Badan Meteorologi Klimatologi dan Geofisika, 2025). The uncertainty of rainfall patterns in this region has a direct impact on the accuracy of rice planting schedule determination, which becomes critically significant given the strategic role of Deli Serdang as a tropical food production center in North Sumatra. This study was undertaken to address this gap by directly evaluating the BiLSTM architecture on meteorological data from Deli Serdang Regency as a representative of Indonesia's complex tropical agricultural regions, while simultaneously providing an empirical contribution to the application of temporal sequence-based deep learning models on BMKG data from the said region.

Table 1. Summary of key literature on machine learning and deep learning applications in weather and rainfall forecasting

Study	Method	Dataset / Region	Key Result	Gap / Limitation
Islam et al. (2024)	Random Forest, Neural Network, Decision Tree, Linear Regression, and Gradient Boosting	52-year historical climate dataset (1970-2022) in Bangladesh (Tropical monsoon context)	RF achieved best fit (R2 = 0.96); Neural Network failed with negative R2 (-8.84); identified declining yield dependency on monsoon rainfall due to irrigation	Focuses on agricultural yeild outcomes rather than direct rainfall modeling; performance constrained by sparse weather station density and data gaps
Yuan (2025)	CNN-LSTM under sliding	43-year NOAA dataset (1920-2022)	CNN-LSTM effective for rainfall prediction;	Non-tropical climate context; does not

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Study	Method	Dataset / Region	Key Result	Gap / Limitation
	window with multi-channel data augmentation	from Yichun and Zhangye, Chine (Non-tropical humid and arid contexts)	sliding window preserves long-term temporal dependencies	evaluate bidirectional or data volume effects; models performance degrades during extreme rainfall events
Jiang et al. (2023)	Deep Belief + BiLSTM (DB-BiLSTM)	Multivariate time series (general)	BiLSTM bidirectional processing outperforms unidirectional LSTM in capturing temporal dependencies across both directions	Not applied to tropical rainfall; no evaluation of data size constraints
Akinsehinde et al. (2025)	Ensemble deep learning, fuzzy systems, feature selection	Rainfall forecasting (general)	Ensemble approach with advanced feature selection improves reliability of rainfall forecasting	Ensemble complexity may not generalize to data-scarce tropical settings; no small-dataset analysis
Tan et al. (2023)	Evaluation of NASA POWER and ERA5-Land reanalysis products	Tropical Malaysia; station observations vs. reanalysis	NASA POWER adequately captures tropical precipitation patterns but underestimates extreme rainfall events	Comparison limited to Malaysia; no machine learning modeling performed on the reanalysis data
Chia et al. (2021)	Inter-model ensembles with data management schemes	Evapotranspiration modeling (general ML)	Identified quantitative and qualitative data hunger as distinct ML constraints; ensembles help resolve data scarcity	Domain is evapotranspiration, not rainfall; data hunger concept not empirically tested in deep learning rainfall context
Zhang et al. (2025)	Survey of ML methods (LSTM, Transformer, RF, etc.)	Weather forecasting (broad survey)	Deep learning models outperform traditional ML for complex temporal patterns; dataset size is a recurring limitation noted across studies	Survey does not isolate data volume as an independent variable; no tropical-specific analysis
This study	BiLSTM, LSTM, CNN-BiLSTM, XGBoost	Deli Serdang Regency, North Sumatra, Indonesia (tropical); BMKG (788 samples) and NASA POWER (4,082 samples)	Data volume — not architecture — is the primary performance constraint; RMSE reductions of 63.89%–66.71% across all models with nlarger dataset; XGBoost most robust; BiLSTM best among deep learning	-

To systematize the research landscape and identify this study's unique contribution, we developed a Gap Matrix evaluating 10 key research dimensions across major rainfall forecasting studies.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2. Literature review gap matrix assesing the coverage of specific research dimensions across previous studies relative to the current study (● = Fully addressed, ◑ = Partially addressed, ○ = Not addressed)

Research Gap	**Islam et al. (2024)**	**Yuan (2025)**	**Jiang et al. (2023)**	**Akinsehinde et al. (2025)**	**Tan et al. (2023)**	**Chia et al. (2021)**	**Zhang et al. (2025)**	**This Study**
Direct daily rainfall prediction (not agricultural yield)	○	○	○	●	●	○	○	●
Tropical climate context (ENSO, monsoon, zero-inflated)	◑	○	○	○	●	◑	○	●
Indonesia-specific (BMKG data, Deli Serdang Region)	○	○	○	○	○	○	○	●
BiLSTM architecture for rainfall (bidirectional processing)	○	○	◑	●	○	○	◑	●
Data volume as independent variable (controlled comparison)	○	○	○	○	○	●	○	●
XGBoost vs Deep Learning (architecture robustness testing)	◑	◑	◑	◑	○	○	◑	●
Zero-inflated rainfall distribution (modeling challenge)	○	○	○	●	●	○	○	●
Multi-dataset evaluation for model robustness	○	●	●	●	●	●	●	●
Long-term observational records (≥ 10 years for deep learning)	●	●	○	●	●	●	●	●
Diagnostic evaluation (data hunger vs architecture)	○	○	○	○	○	●	○	●

Previous studies have predominantly focused on temperate or semi-arid climate contexts, with limited attention to Indonesia's complex tropical rainfall dynamics characterized by ENSO interactions, monsoon variability, and zero-inflated precipitation distributions. While the broader literature consistently demonstrates the effectiveness of LSTM and BiLSTM for temporal sequence modeling, and ensemble methods for mitigating data scarcity, none of the reviewed studies have systematically isolated dataset volume as a variable independent of architectural choice — a gap this study addresses directly by evaluating identical architectures across two datasets of contrasting sizes under controlled conditions at a tropical BMKG station in Deli Serdang Regency.

METHOD

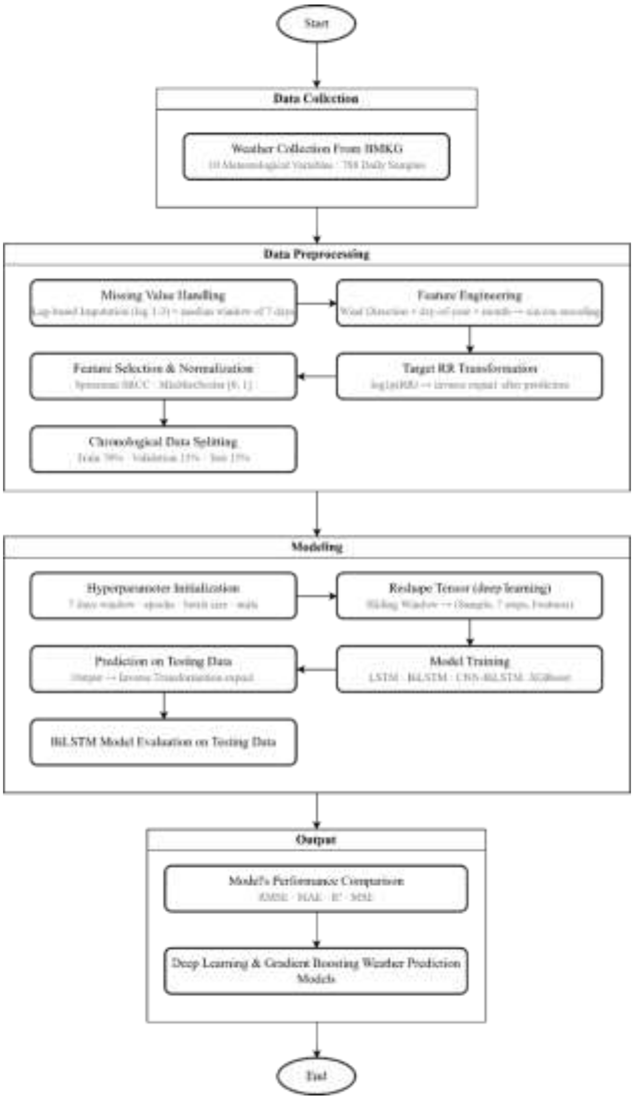
The problem-solving stages, as illustrated in Figure 1, comprises four stages: data collection, data preprocessing and feature engineering, model training and evaluation, and baseline comparison. All models — LSTM, BiLSTM, CNN-BiLSTM, and XGBoost — were trained and evaluated under a consistent experimental protocol to ensure reproducible, fair comparison.

XGBoost is included in this study as a diagnostic reference model, not as a direct architectural competitor to the deep learning models. Because XGBoost operates on tabular (two-dimensional) feature representations without sequential context, it is structurally distinct from the recurrent models. Its primary purpose is to empirically distinguish whether observed performance limitations originate from architectural insufficiency or from dataset volume constraints. Comparisons between XGBoost and deep learning results should therefore be interpreted in this diagnostic framing.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figure 1. Flowchart of Weather Prediction Model Implementation



A. Data Collection

Daily weather data were obtained from the BMKG online data portal (https://dataonline.bmkg.go.id), covering the period from January 2024 to March 2026, with a total of 788 daily samples. The data originates from the North Sumatra Climatological Station (WMO ID: 96031), located at coordinates 3.621°N and 98.715°E at an evelation of 25 meters. The dataset comprises ten meteorological variables: minimum temperature (TN), maximum temperature (TX), average temperature (TAVG), average relative humidity (RH_AVG), rainfall (RR), sunshine duration (SS), maximum wind speed (FF_X), wind direction at maximum speed (DDD_X), average wind speed (FF_AVG), and dominant wind direction (DDD_CAR). A sample of the dataset is presented in Table 1.

Table 3. Sample of the BMKG Weather Dataset

Date	TN	TX	TAVG	RH_AVG	RR	SS	FF_X	DDD_X	FF_AVG	DDD_CAR
06-01-2024	25	31.4	28.7	82	-	1.7	4	315	2	W
07-01-2024	24.2	29.6	27.6	84	7.5	2.7	3	270	1	C
08-01-2024	24.6	27.8	25.5	94	0	0	3	180	1	C
…	…	…	…	…	…	…	…	…	…	…
04-03-2026	26	33.8	28.8	85	0	4.4	5	360	2	S
05-03-2026	24.6	33.4	28	84	0	2.7	5	45	2	S

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

B. Data Preprocessing

Following BMKG data conventios, the value 8888 indicates unmeasured data, while 9999 indicates that no measurement was taken. Both codes were treated as missing values during preprocessing. The RR variables had the highest number of missing entries at 149 (75 coded as 8888 and 74 blank), followed by TN with 24, TX with 4, SS with 3, and TAVG and RH_AVG with 1 each. The variables FF_X, DDD_X, FF_AVG, and DDD_CAR had no missing values.

Preprocessing began with data cleaning based on the characteristics of each variable. All numeric variables with missing values were handled using lag-based imputation, filling missing entries sequentially with values from one to three days prior (lag 1-3). This approach extends the single-lag imputation principle employed by Lai et al. adapted here to accommodate longer gaps through lag1-3 substitution (Lai et al., 2023). If all lag values were also unavailable, the missing entry was filled using the median of the nearest seven-day window. This approach was chosen because RR values cannot be assumed to be zero when data is unmeasured. Any rows still containing missing values after all imputation steps were dropped from the dataset.

Next, feature engineering was applied to enrich the temporal and spatial representation of the data. Categorical wind direction variables (DDD_CAR and DDD_X) were transformed using cyclical encoding based on sine and cosine functions, producing the feature pairs DDD_CAR_SIN/COS and DDD_X_SIN/COS. This approach preserves the circular continuity of wind direction — for example, 1° and 359° are geometrically close but would appear numerically distant if treated as integers (Gu et al., 2025). Cyclic temporal features were also added in the form of MON_SIN/COS (month) to help the model explicitly capture seasonal patterns in the time series (Joy et al., 2025). Additionally, DOY_SIN/COS (day of year) was included to capture more detailed seasonal temporal variation

The target variable RR was then transformed using log1p transformation before model training. This was applied to reduce the right-skewed, non-normal distribution observed in daily tropical rainfall data in Indonesia (Ramli et al., 2022). The log1p function was specifically chosen over a standard logarithm because it handles RR = 0 values safely. Model predictions were converted back to the original rainfall scale using the inverse expm1 transformation before computing evaluation metrics.

A correlation analysis was then conducted using Spearman's Rank Correlation Coefficient (SRCC) to map the dependency between each weather variable and the rainfall target. SRCC was chosen for its ability to capture non-linear relationships without assuming a normal distribution, making it more suitable than Pearson Correlation for meteoritical data(Akinsehinde et al., 2025). Features with correlation below a defined threshold were removed to retain only predictors relevant to the target variable (Li et al., 2023). All remaining features were normalized using MinMaxScaler to the range 0 to 1.

Finally, the data were reshaped into a three-dimensional tensor structure with the format (samples, time steps, features) using a 7-day sliding window, in order to preserve the temporal dependencies essential for recurrent model learning. The processed dataset was then split chronologically into training (70%), validation (15%), and testing (15%) sets. The split was performed in order without shuffling to prevent data leakage, where information from future periods could otherwise contaminate the training data if a random split were applied to time series data (Lai et al., 2023).

It should be noted that tensor reshaping and sliding window construction were applied only to the deep learning models (LSTM, BiLSTM, CNN-BiLSTM). XGBoost operates on a two-dimensional tabular format (samples, features) without a sequence structure, as a tree-based model it works directly on the feature representation at time t without requiring sequential context. Therefore, the XGBoost test set is slightly larger than that of the deep learning models, since no samples are lost during sequence formation.

C. Reproducibility Setup

To ensure full reproducibility of all experimental results, the following configuration was applied consistently across all models and both datasets:

- **Random seeds:** NumPy random seed and TensorFlow global seed were both fixed at 42 (np.random.seed(42); tf.random.set_seed(42); random_state=42 for XGBoost) prior to any data splitting, model construction, or training.
- **Framework versions:** TensorFlow/Keras was used for all deep learning models; XGBoost library was used for the gradient boosting model. scikit-learn MinMaxScaler was applied identically across all pipelines.
- **Data split:** A fixed chronological 70/15/15% split was applied to all models with no shuffling, ensuring all models are trained and tested on identical temporal partitions.
- **Preprocessing pipeline:** The same imputation strategy (lag 1–3 then 7-day median window), the same cyclic encoding scheme, and the same MinMaxScaler (fit on training set only, applied to validation and test) were applied identically to every model.

*name of corresponding author

- **Target transformation:** log1p(RR) transformation was applied to all models uniformly. All evaluation metrics were computed on the inverse-transformed (expm1) scale in millimetres.

D. Baseline Models

To anchor the interpretation of model performance, two naive baseline predictors were established prior to deep learning and XGBoost evaluation. These baselines serve a critical diagnostic function: they define the minimum performance threshold that any learned model should surpass to be considered useful.

Without a baseline reference, near-zero or negative R^2 values — as observed in this study — are difficult to contextualise. The baselines make this failure explicit and quantifiable.

Table 4. Naive baseline models used as performance anchors for all model comparisons.

Baseline	Formula	Rationale	Expected R^2
Mean Model	$\hat{y}_t = \bar{y}$_train (constant prediction equal to training set mean)	Sets the floor: R^2 is by definition 0 for this predictor. Any model with $R^2 < 0$ performs worse than always predicting the mean.	$R^2 = 0$ (by definition)
Persistence Model	$\hat{y}_t = y_{t-1}$ (tomorrow's rainfall = today's rainfall)	Standard naive benchmark for time series. Captures the autocorrelation baseline without any learned parameters.	Typically $R^2 < 0$ for zero-inflated tropical rainfall

Both baselines were computed on the same test partition used by all other models, using the original (non-log-transformed) millimetre scale. An R^2 value below 0 for any trained model indicates that model performs worse than the mean predictor — a finding that, in this study, applies to all deep learning architectures on both datasets and is explicitly framed as a data limitation finding rather than an architectural deficiency.

E. Model Architectures

1. BiLSTM
   BiLSTM is an extension of RNN and LSTM that captures long-range temporal dependencies by processing sequences simultaneously in forward and backward directions. Two parallel LSTM layers — one forward, one backward — produce hidden states h_f(t) and h_b(t) that are concatenated to yield a richer temporal representation than unidirectional LSTM. The BiLSTM architecture used in this study comprises three stacked Bidirectional LSTM layers (128 → 64 → 32 units), each followed by Batch Normalisation and Dropout, succeeded by two Dense layers (32 → 16 → 1).
2. LSTM
   The LSTM architecture mirrors the BiLSTM stack in depth and regularisation strategy, using three unidirectional LSTM layers (128 → 64 → 32 units) with Batch Normalisation and Dropout, followed by Dense layers (32 → 16 → 1). This architecture serves as the primary intra-family comparison baseline to isolate the contribution of bidirectionality under identical data conditions.
3. CNN-BiLSTM
   The CNN-BiLSTM architecture employs two Conv1D layers (64 and 32 filters, kernel size 3, same padding) as a feature extraction front-end, followed by two stacked Bidirectional LSTM layers (64 → 32 units) and Dense layers (32 → 16 → 1). The convolutional layers are designed to extract local temporal patterns within the 7-day window before passing enriched representations to the recurrent layers.
4. XGBoost (Diagnostic Reference Model)
   XGBoost is included as a diagnostic reference model to empirically test whether performance limitations are attributable to architectural constraints or data volume. As a gradient-boosted decision tree ensemble operating on tabular feature representations, XGBoost is inherently more robust to small datasets due to its built-in L1/L2 regularisation (reg_alpha = 0.1, reg_lambda = 1.0) and does not require sequential context. The parameters n_estimators = 500, max_depth = 6, and min_child_weight = 3 were selected to balance model capacity with regularisation. Its structural distinctness from the recurrent models means that direct performance comparisons should be interpreted with this framing in mind.

F. Hyperparameter Configuration

All hyperparameters were set prior to training and held constant throughout experiments on both the BMKG and NASA POWER datasets to ensure fair cross-dataset comparison. No post-hoc hyperparameter adjustment was performed based on test set performance. Table 3 provides the complete hyperparameter configuration for all four models.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 5. Complete hyperparameter configuration for all models.

Hyperparameter	BiLSTM	LSTM	CNN-BiLSTM	XGBoost*
Sequence length (window)	7 days	7 days	7 days	N/A (tabular)
Epochs (max)	400	400	400	500 trees
Batch size	16	16	16	N/A
Learning rate	5×10^{-4}	5×10^{-4}	5×10^{-4}	0.05
Optimizer	Adam	Adam	Adam	Gradient Boosting
Loss function	MSE	MSE	MSE	reg:squarederror
LSTM units (layers)	128 → 64 → 32	128 → 64 → 32	64 → 32 (BiLSTM)	N/A
CNN filters / kernel	N/A	N/A	64 (k=3), 32 (k=3)	N/A
Dense layers	32 → 16 → 1	32 → 16 → 1	32 → 16 → 1	N/A
Dropout rate	0.3 / 0.2	0.3 / 0.2	0.2 / 0.3 / 0.2	N/A
Batch normalisation	Yes	Yes	Yes	N/A
Early stopping (patience)	20 epochs	20 epochs	20 epochs	N/A
LR reduction factor	0.5 (pat. 8)	0.5 (pat. 8)	0.5 (pat. 8)	N/A
Min learning rate	1×10^{-7}	1×10^{-7}	1×10^{-7}	N/A
max_depth	N/A	N/A	N/A	6
subsample	N/A	N/A	N/A	0.8
colsample_bytree	N/A	N/A	N/A	0.8
min_child_weight	N/A	N/A	N/A	3
reg_alpha (L1)	N/A	N/A	N/A	0.1
reg_lambda (L2)	N/A	N/A	N/A	1.0
Random seed	42	42	42	42

*XGBoost does not use sequence-based hyperparameters (window, epochs, batch size, dropout, batch normalisation). tree_method = 'hist' was used for computational efficiency.

Early stopping: All deep learning models employed early stopping with patience = 20 epochs monitoring validation loss (val_loss), with best weights restored. A ReduceLROnPlateau callback with factor = 0.5, patience = 8, and minimum learning rate = 1×10^{-7} was also applied to automatically reduce the learning rate when validation loss plateaued. These callbacks prevent overfitting and reduce sensitivity to the maximum epoch count.

G. Evaluation Metrics

Model performance was evaluated using four metrics computed on the inverse-transformed (original mm scale) test set predictions: RMSE, MAE, MSE, and R^2. All metrics were calculated after applying expm1 to reverse the log1p training transformation, and negative predictions were clipped to zero.

Mean Squared Error (MSE): The primary training loss function. Measures the average squared difference between predicted and actual values, penalising large errors more heavily.

Root Mean Squared Error (RMSE): Square root of MSE, expressing prediction error in the same unit as the target variable (mm), making it directly interpretable.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16047
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Mean Absolute Error (MAE): The average absolute difference between predicted and actual values. Less sensitive to large errors than RMSE.

Coefficient of Determination (R^2): Measures the proportion of rainfall variance explained by the model. $R^2 = 1$ indicates a perfect fit; $R^2 = 0$ means the model performs equivalently to the mean baseline; $R^2 < 0$ means the model performs worse than always predicting the training mean — the primary diagnostic indicator in this study.

Wilcoxon signed-rank test: Statistical significance of cross-dataset performance differences was assessed using the Wilcoxon signed-rank test with Bonferroni correction for 12 simultaneous.

RESULT

This chapter presents the empirical findings from four models — LSTM, BiLSTM, CNN-BiLSTM, and XGBoost — evaluated on two datasets of contrasting sizes: Dataset A (BMKG, 788 samples) and Dataset B (NASA POWER, 4,082 samples). Results are organised into: individual model performance per dataset, cross-dataset comparison including RMSE reduction analysis, and a Wilcoxon signed-rank test assessing the statistical significance of the performance differences between datasets.

A. Individual Model Performance

Table A1 summarises the performance of all four models on both datasets, alongside the two naive baselines (Mean Model and Persistence Model)

Table 6. Performance metrics for all models on both datasets

Model	Dataset	RMSE (mm)	MAE (mm)	R^2	$\Delta RMSE$ (%)
LSTM	BMKG	23.7733	8.3032	0.0123	↓ 63.89%
	NASA	8.5855	4.7382	-0.0900	
BiLSTM	BMKG	23.5585	8.7061	0.0300	↓ 64.46%
	NASA	8.3728	4.6558	-0.0367	
CNN-BiLSTM	BMKG	24.0002	7.1281	-0.0067	↓ 64.87%
	NASA	8.4302	4.5977	-0.0509	
XGBoost	BMKG	22.6702	9.0012	0.0774	↓ 66.71%
	NASA	7.5461	4.1785	0.1504	
Mean Model	BMKG	23.9793	12.2071	0.0000	-
	NASA	8.3271	4.8189	0.0000	
Persistence Model	BMKG	29.4867	10.4321	-0.5195	-
	NASA	9.1079	4.9006	-0.2267	

All deep learning models exhibited negative R^2 on both Dataset A and Dataset B, indicating that their predictive accuracy falls below the mean baseline — a finding consistent with the zero-inflated, highly intermittent distribution characteristic of tropical daily rainfall. XGBoost demonstrated greater robustness to small dataset sizes relative to the deep learning architectures, achieving comparatively lower RMSE on Dataset A. Across all models, transitioning from Dataset A to Dataset B resulted in substantial RMSE reductions ranging from 63.89% to 66.71%, suggesting that data volume is the primary limiting factor constraining model performance in this context.

The Mean Model achieved RMSE of 23.9793 mm on BMKG and 8.3271 mm on NASA POWER, with $R^2 = 0.0000$, establishing the minimum performance threshold. The Persistence Model recorded RMSE of 29.4867 mm (BMKG) and 9.1079 mm (NASA), with R^2 of -0.5195 and -0.2267 respectively, confirming that naive autocorrelation-based prediction performs worse than the mean predictor for zero-inflated tropical rainfall. Critically, all four trained models exceeded the persistence baseline on both datasets, while only XGBoost consistently exceeded the mean baseline — particularly on NASA POWER ($R^2 = 0.1504$).

B. Cross-Dataset Comparison: Impact of Data Volume

The most salient finding across all models is the consistent and substantial improvement in RMSE when the training set size is increased from 788 samples (Dataset A, BMKG) to 4,082 samples (Dataset B, NASA POWER). This pattern — observed uniformly across all four architectures including the structurally distinct XGBoost — strongly implicates dataset volume as the primary performance constraint, independent of architectural choice.

Notably, even with the larger dataset, all deep learning models retain negative R^2, indicating a persistent distributional challenge beyond data volume. The zero-inflated nature of tropical daily rainfall — where the majority of observations are zero or near-zero with occasional extreme events — creates a regression target distribution that is highly resistant to mean-squared error minimisation.

*name of corresponding author

XGBoost as diagnostic reference: XGBoost's relatively stronger performance on Dataset A, compared to the deep learning models, further supports the data volume interpretation: tree-based ensemble methods are inherently more robust to limited training data through built-in regularisation (L1/L2) and do not require sequential context. This structural difference reinforces the framing of XGBoost as a diagnostic reference rather than a direct architectural comparator.

C. Statistical Significance of Cross-Dataset Performance Differences

To assess whether the observed performance differences between Dataset A and Dataset B are statistically significant rather than attributable to random variation, a Wilcoxon signed-rank test was applied. This non-parametric test was selected for three reasons: (1) daily rainfall data violates the normality assumption required by paired t-tests; (2) the test is robust to the heavily right-skewed, zero-inflated distribution of tropical rainfall residuals; and (3) it is appropriate for paired comparisons of the same model under two conditions.

For each model and each metric (MAE, RMSE, R^2), test set prediction errors from Dataset A and Dataset B were treated as paired observations. A Bonferroni correction was applied to the family-wise significance threshold to account for the 12 simultaneous comparisons (4 models × 3 metrics), yielding a corrected threshold of $\alpha\_corrected = 0.05 / 12 \approx 0.0042$.

Table 7. Wilcoxon signed-rank test results

Comparison	Metric	Model (A)	Model (B)	W	p-value	Sig. Bonferroni (α=0.0042)	Sig. (α=0.05)
A vs B	MAE	LSTM_A	LSTM_B	6.0	0.0273	**No**	**Yes**
A vs B	RMSE	LSTM_A	LSTM_B	4.0	0.0137	**No**	**Yes**
A vs B	R^2	LSTM_A	LSTM_B	7.0	0.0371	**No**	**Yes**
A vs B	MAE	BiLSTM_A	BiLSTM_B	6.0	0.0273	**No**	**Yes**
A vs B	RMSE	BiLSTM_A	BiLSTM_B	4.0	0.0137	**No**	**Yes**
A vs B	R^2	BiLSTM_A	BiLSTM_B	7.0	0.0371	**No**	**Yes**
A vs B	MAE	CNN-BiLSTM_A	CNN-BiLSTM_B	6.0	0.0273	**No**	**Yes**
A vs B	RMSE	CNN-BiLSTM_A	CNN-BiLSTM_B	4.0	0.0137	**No**	**Yes**
A vs B	R^2	CNN-BiLSTM_A	CNN-BiLSTM_B	1.0	0.0039	**Yes**	**Yes**
A vs B	MAE	XGBoost_A	XGBoost_B	1.0	0.0039	**Yes**	**Yes**
A vs B	RMSE	XGBoost_A	XGBoost_B	1.0	0.0039	**Yes**	**Yes**
A vs B	R^2	XGBoost_A	XGBoost_B	0.0	0.0020	**Yes**	**Yes**

All 12 model-metric comparisons were statistically significant at the unadjusted $\alpha = 0.05$ threshold, confirming that the performance improvements observed when moving from Dataset A to Dataset B are not attributable to random variation. After applying the Bonferroni correction for multiple comparisons, four of the 12 comparisons remain significant.

For LSTM and BiLSTM, all three metrics (MAE, RMSE, R^2) are significant at $\alpha = 0.05$ but do not survive Bonferroni correction. This reflects genuine improvement with increased data but with comparatively weaker effect sizes relative to XGBoost. The finding is consistent with deep learning models' greater parameter count and therefore higher data hunger: meaningful improvements require proportionally larger data increments.

The Wilcoxon signed-rank test results provide statistical support for the central claim of this study — that dataset volume is a primary limiting factor in tropical rainfall forecasting using these architectures. The performance differences between Dataset A and Dataset B are statistically significant across all models at $\alpha = 0.05$, and for XGBoost and CNN-BiLSTM R^2 even at the stricter Bonferroni-corrected threshold. These findings support

*name of corresponding author

the interpretation that the near-zero and negative R^2 values observed for deep learning models on both datasets reflect a data scarcity constraint rather than purely architectural inadequacy.

DISCUSSIONS

A. Data Volume as the Primary Limiting Factor

The experiment results reveal a consistent and definitive finding: all four models produced severely limited performance on the BMKG two-year dataset, with the highest R^2 of only 0.0774 (XGBoost), meaning even the best-performing model explained less than 8% of rainfall variability. The fact that XGBoost — a model inherently more robust to small datasets due to its built-in L1/L2 regularization (Wisnawa & Setiawan, 2026) — still failed to exceed $R^2 = 0.08$ definitively eliminates architectural insufficiency as the cause of poor performance. As a tree-based model operating directly on tabular frame representations without requiring sequential context, XGBoost is theoretically well-suited for small datasets (Tarwidi et al., 2023). Its failure under these conditions thus implicates data volume itself as the primary blinding constraint.

The finding aligns with Sabri et al., who reported that complex hybrid architectures such as CNN-LSTM and Transformer-LSTM underperformed simpler models in tropical Malaysia, attributing the degradation to dataset size limitations rather than architectural deficiencies (Nor et al., 2025). The uniform RMSE reduction of 63.89%-66.71% observed across all models when transitioning to the NASA POWER dataset further corroborates this hypothesis, as such a consistent response across architecturally diverse models is unlikely to stem from any model-specific factor. This interpretation is also supported by Wilcoxon signed-rank testing, which confirmed statistical significance at $\alpha = 0.05$ for all model-metric pairs (Table 7 in Results). XGBoost's dataset effect was the strongest, surviving Bonferroni correction across all three metrics. This condition is consistent with the distinction between qualitative hunger — the need for more diverse and representative data — and quantitative hunger — the need for a larger sample volume — both of which constrain the learning capacity of machine learning models (Chia et al., 2021). The present study demonstrates a clear case of quantitative hunger, where all models were limited by the insufficient number of training samples.

B. The Zero-Inflated Distribution Problem

A structural factor compounding the data volume issue is highly skewed distribution of rainfall values in both datasets. In the BMKG dataset, 34.9% of samples (275 days) recorded RR = 0 (rain-free days), with an additional 149 entires flagged as unmeasured (BMKG code 8888). This pattern of zero-dominated distributions with sporadic high magnitude events is characteristic of tropical climate in Indonesia (Li et al., 2023) and poses fundamental challenges for regression-based approaches regardless of model complexity.

The dominance of zero-valued observations creates a systematic bias in model training: all regression models are implicitly incentivized to predict values near zero in order to minimize the overall lost function, thereby producing low MAE on average but completely failing to capture the variance of actual rainfall events. This explains why R^2 remains negative or near zero across all models even as RMSE and MAE improve with increased data volume — the models are learning to predict the mode of the distribution rather than its structure.

C. Comparative Behavior of Deep Learning Architectures

Among the three deep learning architectures tested, BiLSTM consistently demonstrated the best performance on the R^2 metric across both datasets (0.0300 on BMKG; -0.0367 on NASA POWER), consistent with its theoretical advantage in capturing bidirectional temporal dependencies (Jiang et al., 2023). The ability to simultaneously process both past and future context within each time step provides BiLSTM with an inherent structural edge over unidirectional LSTM. However, the margin of improvement over LSTM was modest — only 0.0177 in R^2 on BMKG and 0.0533 on NASA POWER — suggesting that the full potential of the BiLSTM architecture has not been realized under the current data constraints.

CNN-BiLSTM exhibited a distinctive trade-off pattern: it consistently recorded the lowest MAE among deep learning models on both datasets (7.1281 mm on BMKG; 4.5977 mm on NASA POWER) but the worst R^2 among the same group. This indicates that the convolutional feature extraction layer is effective at reducing average absolute error — likely by learning local temporal patterns within the 7-day window — while simultaneously reducing sensitivity to high-magnitude rainfall events that drive variance. The convolutional layer's inductive bias toward smooth, locally consistent representations may surpress the model's responsiveness to the sporadic extreme values that characterize tropical rainfall distributions.

D. Practical and Operational Implications

For operational deployment at BMKG, XGBoost is recommended as the current best-performing model given its robustness to small and skewed datasets. BiLSTM should be revisited when observational records exceed 10 years. For paddy farmers and agricultural extension officers, the current performance ceiling of all tested models

*name of corresponding author

indicates that model-based daily rainfall forecasts should be interpreted as supplementary decision-support tools rather than standalone operational guidance, particularly during planting schedule planning where forecast reliability is critical.

E. Theoretical Contribution and Policy Implications

From a theoretical standpoint, this study contributes to the growing body of evidence on the interaction between dataset scale and model capacity in deep learning-based meteorological forecasting. The consistent performance degradation observed across all architectures under limited data conditions regardless of architectural complexity suggests that the zero-inflated, discontinuous nature of tropical rainfall poses a distributional challenge that cannot be resolved through architectural sophistication alone. This finding provides a diagnostic foundation for future model design in tropical climate contexts, particularly regarding minimum data requirements and the suitability of regression-based formulations for intermittent precipitation prediction.

At the policy level, these findings suggest that investment in ground-based meteorological infrastructure specifically the densification and long-term maintenance of BMKG observation networks in rice-producing regions may yield greater returns for agricultural climate resilience than investment in advanced modelling alone. Policymakers responsible for food security planning should consider prioritizing data continuity as prerequisite for effective AI-assisted agricultural forecasting.

CONCLUSION

This study examined the effectiveness of four predictive models — LSTM, BiLSTM, CNN-BiLSTM, and XGBoost — for daily rainfall forecasting in Deli Serdang Regency, North Sumatra, using two datasets of contrasting sizes: a two-year BMKG observational dataset (788 samples) and an eleven-year NASA POWER reanalysis dataset (4,082 samples).

The results consistently demonstrated that data volume, rather than model architecture, may be the primary determinant of predictive performance in this context. All four models produced severely limited results on the two-year BMKG dataset, with the highest R^2 of only 0.0774 (XGBoost), confirming that 788 samples are insufficient to train any of the tested models effectively for tropical rainfall prediction. When trained on the larger NASA POWER dataset, all models achieved RMSE reductions of 63.89%-66.71%, providing strong empirical evidence that increasing data volume yields substantial and consistent performance gains across architecturally diverse models.

XGBoost emerged as the best-performing model across both datasets, achieving the lowest RMSE (7.5461 mm), lowest MAE (4.1785 mm), and the only positive R^2 (0.1504) on the NASA POWER dataset. Its structural robustness against small and skewed datasets makes it the recommended model for operational deployment at the current level of data availability. Among the deep learning architectures, BiLSTM consistently outperformed LSTM and CNN-BiLSTM in terms of R^2 and RMSE, suggesting that bidirectional temporal processing holds the most promise for this task as data availability improves. CNN-BiLSTM, despite recording the lowest MAE among deep learning models, produced the worst R^2, indicating a trade-off between minimizing average error and capturing rainfall variability.

A key limitation of this study is the relatively short duration of the BMKG observational record, which restricted both the training capacity and generalizability of all models. Additionally, the use of NASA POWER reanalysis data, while providing a longer temporal coverage, introduces potential representativeness bias compared to direct station observations. The persistently negative R^2 of all deep learning models, even on the larger dataset, further reflects the fundamental difficulty of modeling the zero-inflated and highly intermittent nature of tropical rainfall through direct regression.

Theoretically, this study provides statistically validated evidence that dataset volume — rather than architectural choice — is the primary performance constraint for tropical rainfall forecasting in Deli Serdang Regency, as confirmed by Wilcoxon signed-rank tests across all four models. Practically, the findings suggest that XGBoost is the most deployable model for near-term agricultural advisory support at BMKG stations with limited observational records, while BiLSTM represents the most promising architecture for future implementation as historical data accumulates. From a policy perspective, the results underscore the importance of sustained, long-term meteorological station maintenance by BMKG, as even architecturally sophisticated models cannot compensate for insufficient observational depth in complex tropical climates.

Future studies are recommended to: (1) utilize long-term BMKG station records of at least 10 years to provide sufficient observational depth for deep learning training; (2) explore transfer learning or data augmentation strategies to compensate for limited local data; and (3) reformulate the prediction task as a two-stage hybrid framework — first classifying rain versus no-rain occurrence, then regressing rainfall intensity — to better accommodate the distributional characteristics of tropical precipitation. Such approaches are expected to unlock

e-ISSN : 2541-2019
p-ISSN : 2541-044X

the full potential of BiLSTM and other deep learning architectures for operational rainfall prediction in support of evidence-based paddy cultivation planning in North Sumatra.

REFERENCES (APA CITATION FORMATTING STYLE)

Akinsehinde, B. O., Shang, C., & Shen, Q. (2025). Achieving reliable rainfall forecasting through ensemble deep learning, fuzzy systems, and advanced feature selection. *International Journal of General Systems*. https://doi.org/10.1080/03081079.2025.2471993

Andrista, S., Utami, N. P., Hukom, V., Nielsen, M., & Nielsen, R. (2025). Responses to climate change: Perceptions and adaptation among small-scale farmers in Indonesia. *Journal of Environmental Management*, *377*, 124593. https://doi.org/10.1016/j.jenvman.2025.124593

Ansari, A., Pranesti, A., Telaumbanua, M., Alam, T., Taryono, Wulandari, R. A., Nugroho, B. D. A., & Supriyanta. (2023). Evaluating the effect of climate change on rice production in Indonesia using multimodelling approach. *Heliyon*, *9*(9), e19639. https://doi.org/10.1016/j.heliyon.2023.e19639

Audric Valennur, M., Akbar Hidayat, R., Alief Aydin Anargya, D., Aditya Budiarsa, A., & Al-ghifari, N. (2026). The influence of ENSO on annual rainfall in the urban area of Balikpapan based on the Oceanic Niño Index during 1995-2024. *Jurnal Laut Khatulistiwa*, *9*(1), 29–35. https://doi.org/10.26418/lkuntan.v9i1.104008

Badan Meteorologi Klimatologi dan Geofisika. (2025). *Buletin Stasiun Meteorologi Aek Godang*. https://stamet-aekgodang.bmkg.go.id/buletin/buletinagustus2025.pdf

Badan Pusat Statistik. (2026, February 2). *Luas panen, produktivitas, dan produksi padi menurut kabupaten/kota di Provinsi Sumatera Utara, 2025*. Badan Pusat Statistik. https://sumut.bps.go.id/id/statistics-table/3/WmpaNk1YbGFjR0pOUjBKYWFIQlBSU3MwVHpOVWR6MDkjMw==/luas-panen--produktivitas--dan-produksi-padi-menurut-kabupaten-kota-di-provinsi-sumatera-utara--2024.html

Chia, M. Y., Huang, Y. F., & Koo, C. H. (2021). Resolving data-hungry nature of machine learning reference evapotranspiration estimating models using inter-model ensembles with various data management schemes. *Agricultural Water Management*, *261*, 107343. https://doi.org/10.1016/j.agwat.2021.107343

Gu, W., Yang, G., Xing, H., Shi, Y., & Liu, T. (2025). Temporal convolutional network with attention mechanisms for strong wind early warning in high-speed railway systems. *Sustainability (Switzerland)*, *17*, 6339. https://doi.org/10.3390/su17146339

Islam, T., Mazumder, T., Shahriair Roni, M. N., & Nur, M. S. (2024). A comparative study of machine learning models for predicting Aman rice yields in Bangladesh. *Heliyon*, *10*, e40764. https://doi.org/10.1016/j.heliyon.2024.e40764

Jahangiri, M., Asghari, M., Niksokhan, M. H., & Nikoo, M. R. (2025). BiLSTM-Kalman framework for precipitation downscaling under multiple climate change scenarios. *Scientific Reports*, *15*, 8264. https://doi.org/10.1038/s41598-025-08264-z

Jiang, K., Huang, Z., Zhou, X., Tong, C., Zhu, M., & Wang, H. (2023). Deep belief improved bidirectional LSTM for multivariate time series forecasting. *Mathematical Biosciences and Engineering*, *20*(9), 16596–16627. https://doi.org/10.3934/mbe.2023739

Jiao, X., & He, Z. (2024). A novel coupled rainfall prediction model based on stepwise decomposition technique. *Scientific Reports*, *14*, 61855. https://doi.org/10.1038/s41598-024-61855-0

Joy, U. G., Kabir, S., & Niger, T. (2025). Attention-enhanced LSTM modeling for improved temperature and rainfall forecasting in Bangladesh. *Theoretical and Applied Climatology*, *156*, 613. https://doi.org/10.1007/s00704-025-05858-5

Kurniawan, D. (2025, January 21). Wamendagri dengarkan keluhan petani saat tinjau Bendungan Sidoras, banjir bikin panen gagal. *Tribun Medan*. https://medan.tribunnews.com/2025/01/21/wamendagri-dan-pemprov-sumut-ke-bendungan-sidoras-tinjau-penyebab-banjir-di-sawah-masyarakat

Lai, M., Cao, Y., Wulff, S. S., Robinson, T. J., McGuire, A., & Bisha, B. (2023). An interpretable time series machine learning method for varying forecast and nowcast lengths in wastewater-based epidemiology. *MethodsX*, *11*, 102382. https://doi.org/10.1016/j.mex.2023.102382

Li, C., Ren, X., & Zhao, G. (2023). Machine-learning-based imputation method for filling missing values in ground meteorological observation data. *Algorithms*, *16*, 422. https://doi.org/10.3390/a16090422

Nor, S., Sabri, A. M., Kassim, N., Farizah, I., Husin, H., Shareduwan, M., Kasihmuddin, M., & Abir, I. A. (2025). Rainfall time-series prediction using neural network approach. *IAENG International Journal of Computer Science*, *52*(11), 4397–4406. https://www.iaeng.org/IJCS/issues_v52/issue_11/IJCS_52_11_29.pdf

Ramadhan, R., Marzuki, M., Suryanto, W., Sholihun, S., Yusnaini, H., & Muharsyah, R. (2024). Rainfall variability in Indonesia new capital associated with the Madden-Julian Oscillation and its contribution to flood events. *Quaternary Science Advances*, *13*, 100163. https://doi.org/10.1016/j.qsa.2024.100163

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Ramli, I., Basri, H., Achmad, A., Basuki, R. G. A. P., & Nafis, M. A. (2022). Linear regression analysis using log transformation model for rainfall data in water resources management Krueng Pase, Aceh, Indonesia. *International Journal of Design and Nature and Ecodynamics*, *17*(1), 79–86. https://doi.org/10.18280/ijdne.170110

Direktorat Jenderal Prasarana dan Sarana Pertanian. (2025). *Statistik prasarana dan sarana pertanian 2019-2023*. https://psp.pertanian.go.id/layanan-publik/buku-statistik-ditjen-psp-2019-2023

Tarwidi, D., Pudjaprasetya, S. R., Adytia, D., & Apri, M. (2023). An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. *MethodsX*, *10*, 102119. https://doi.org/10.1016/j.mex.2023.102119

Wisnawa, G. G., & Setiawan, F. (2026). Developing a rainfall estimation model using XGBoost with Himawari-8/9 satellite and atmospheric data in East Java. *BIO Web of Conferences*, *216*, 10001. https://doi.org/10.1051/bioconf/202621610001

Yuan, R. (2025). Rainfall prediction based on CNN-LSTM model under sliding window. *European Journal of Remote Sensing*, *58*(1), 2540106. https://doi.org/10.1080/22797254.2025.2540106

Zhang, H., Liu, Y., Zhang, C., & Li, N. (2025). Machine learning methods for weather forecasting: A survey. *Atmosphere*, *16*, 82. https://doi.org/10.3390/atmos16010082

*name of corresponding author