An Adaptive Grey Wolf Optimized Bidirectional LSTM Framework for Flood Risk-Oriented Rainfall Forecasting in Tropical Climate Systems

Yuslena Sari
Department of Information Technology, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
yuzlena@ulm.ac.id (corresponding author)

Nurul Fathanah Mustamin
Department of Information Technology, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
nurul.mustamin@ulm.ac.id

Ahmad Rusyadi
Department of Information Technology, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
lutentiums@gmail.com

Firman Aziz
Computer Science Department, Universitas Pancasakti, Makassar, Indonesia
Firman.aziz@unpacti.ac.id

Received: 13 February 2026 | Revised: 7 April 2026 | Accepted: 17 April 2026

| DOI: https://doi.org/10.48084/etasr.18161

ABSTRACT

This study presents a flood-oriented rainfall forecasting framework that integrates Bidirectional Long Short-Term Memory (BiLSTM) with Grey Wolf Optimizer (GWO) and Shapley Additive Explanations (SHAP)-based interpretability. The proposed approach addresses two key limitations in existing Deep Learning (DL) rainfall models: suboptimal hyperparameter selection and limited model transparency. GWO is employed to optimize network hyperparameters in a stability-aware manner, enhancing convergence behavior and reducing generalization variance. To mitigate the black-box nature of recurrent networks, SHAP analysis is incorporated to quantify feature contributions and ensure physically consistent interpretation of rainfall drivers. Experimental results demonstrate that the optimized GWO–BiLSTM model achieves superior predictive accuracy (R^2 = 0.912) compared to Gated Recurrent Unit (GRU) and non-optimized BiLSTM baselines. In addition to improved regression performance, the model exhibits enhanced sensitivity to extreme rainfall events, reflected by higher recall and lower false negative rates, which are essential for flood early-warning applications. Comparative analysis with recent international studies indicates that the proposed framework uniquely combines metaheuristic optimization, extreme-event evaluation, statistical validation, and explainable modeling within a unified engineering architecture. The results confirm that integrating optimization robustness and interpretable DL significantly strengthens the operational reliability of rainfall prediction systems for hydrological risk mitigation.

Keywords-rainfall prediction; flood forecasting; Bidirectional Long Short-Term Memory (BiLSTM); Grey Wolf Optimizer (GWO); extreme event detection

I. INTRODUCTION

Flood disasters remain one of the most destructive natural hazards worldwide, particularly in tropical and monsoon regions where high-intensity short-duration rainfall frequently triggers flash floods and severe hydrological disruptions. According to recent hydrological assessments, the increasing variability of precipitation patterns due to climate change has intensified the frequency and magnitude of extreme rainfall events, significantly increasing flood risk exposure [1, 2]. Accurate rainfall forecasting therefore plays a crucial role in flood early-warning systems, disaster risk reduction, and infrastructure resilience planning.

Conventional hydrological and Numerical Weather Prediction (NWP) models rely heavily on physical parameterization and complex differential equations to simulate rainfall–runoff processes. Although physically interpretable, these approaches often require extensive calibration, high computational cost, and large-scale meteorological inputs, limiting their adaptability in data-scarce or real-time operational environments [3]. Furthermore, physically based models may struggle to capture nonlinear temporal dependencies during extreme rainfall events.

In recent years, Machine Learning (ML) and Deep Learning (DL) techniques have emerged as powerful alternatives for hydrological time-series forecasting. Systematic reviews confirm that deep neural architectures significantly outperform traditional statistical and conceptual models in rainfall and flood prediction tasks [2, 4, 5]. Among these architectures, Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), demonstrate superior capability in modeling long-term temporal dependencies inherent in hydrometeorological data [6-10].

Recent advancements in artificial intelligence have significantly transformed rainfall forecasting research, particularly through DL architectures capable of modeling nonlinear temporal dependencies in hydrometeorological data [6, 11]. Among these, LSTM networks have been widely adopted due to their ability to capture long-term sequential patterns and mitigate vanishing gradient issues [12-16]. These models have demonstrated superior performance over conventional statistical and conceptual hydrological approaches [17-19].

In addition to LSTM, GRU networks have emerged as computationally efficient alternatives, offering reduced parameter complexity while maintaining competitive predictive performance. Comparative analyses suggest that the relative effectiveness of LSTM and GRU architectures is often dataset-dependent, highlighting the importance of architecture benchmarking in practical forecasting systems [7, 20].

Beyond RNNs, alternative DL approaches such as Temporal Convolutional Networks (TCN) and physics-guided neural networks have been introduced to enhance predictive stability and incorporate hydrological constraints [21-25]. While these methods improve generalization, most existing studies primarily focus on minimizing global error metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), without explicitly addressing performance under extreme rainfall conditions that are critical for flood risk mitigation [26-28].

To enhance model performance, metaheuristic optimization techniques such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Grey Wolf Optimizer (GWO) have been increasingly applied for hyperparameter tuning. Among these, GWO provides a balanced exploration–exploitation mechanism with relatively low computational complexity [17, 18]. However, most optimization-based studies focus solely on improving average prediction accuracy, without incorporating flood-sensitive evaluation metrics such as extreme-event recall or false-negative reduction [29, 30].

Another important limitation in current research is the lack of interpretability. DL models are often criticized as black-box systems, limiting their applicability in operational disaster management [26, 31-34]. Explainable Artificial Intelligence (XAI) techniques, particularly Shapley Additive Explanations (SHAP), have shown potential in quantifying feature contributions and improving model transparency [35-37]. Nevertheless, explainability is typically treated as a post-hoc analysis rather than being integrated into the core modeling framework.

Despite these advances, three major challenges remain.

First, most DL rainfall forecasting studies primarily optimize error-based metrics such as RMSE and MAE without explicitly addressing performance under extreme rainfall conditions, which are critical for flood risk mitigation [3, 38]. Underestimation of peak rainfall events may lead to severe operational consequences in early-warning systems.

Second, hyperparameter selection in deep neural networks is often performed manually or through grid/random search, which may not yield optimal model configurations. Metaheuristic optimization techniques such as PSO, GA, and GWO have been increasingly integrated into hydrological forecasting models to enhance parameter tuning efficiency [20, 29, 39, 40].

Third, DL models are frequently criticized for lacking interpretability, limiting their adoption in operational disaster management. XAI techniques such as SHAP provide a unified framework to quantify feature contributions and improve transparency in prediction models [32, 41]. Such approaches have been used to identify dominant climatic drivers influencing model outputs [42, 43]. Interpretability is particularly important in flood risk applications, where decision-makers require reliable explanations to validate early-warning triggers.

Recent hybrid DL frameworks combining LSTM, GRU, and optimization algorithms demonstrate improved predictive stability and robustness for extreme hydrological conditions [44, 45]. However, limited research simultaneously integrates (i) metaheuristic-based adaptive optimization, (ii) comparative evaluation between BiLSTM and GRU architectures, (iii) flood-risk-oriented performance metrics, and (iv) SHAP-based interpretability analysis within a unified engineering framework.

To address these research gaps, this study proposes a GWO-guided Bidirectional LSTM (GWO–BiLSTM) framework for flood-oriented rainfall forecasting in tropical regions. Unlike conventional accuracy-driven approaches, the proposed method emphasizes:

1. Adaptive hyperparameter optimization using GWO to enhance prediction stability and minimize extreme rainfall underestimation.

2. Comparative architecture evaluation between BiLSTM and GRU to assess temporal learning efficiency.
3. Flood-risk-sensitive metrics, including extreme rainfall recall and false-negative rate analysis.
4. SHAP-based feature importance analysis to provide transparent interpretation of meteorological drivers influencing flood-trigger rainfall events.

By integrating optimization, comparative DL architectures, and explainability within a disaster-oriented evaluation framework, this study offers a practical and engineering-applicable methodology to enhance rainfall forecasting reliability in flood early-warning systems. Unlike previous studies that focus solely on either optimization or DL, this work unifies four key components within a single framework: (i) adaptive metaheuristic optimization, (ii) comparative evaluation of model architectures, (iii) flood-risk-oriented performance metrics, and (iv) SHAP-based interpretability; to the best of our knowledge, no prior study has simultaneously addressed all these aspects within a single rainfall forecasting system.

II. PROPOSED METHODOLOGY

This study proposes a GWO–BiLSTM framework for flood-oriented rainfall forecasting. The proposed methodology integrates multivariate meteorological feature engineering, adaptive hyperparameter optimization, comparative architecture benchmarking (BiLSTM vs. GRU), and SHAP-based interpretability within a unified disaster-sensitive evaluation framework. The overall system workflow is illustrated in Figure 1.

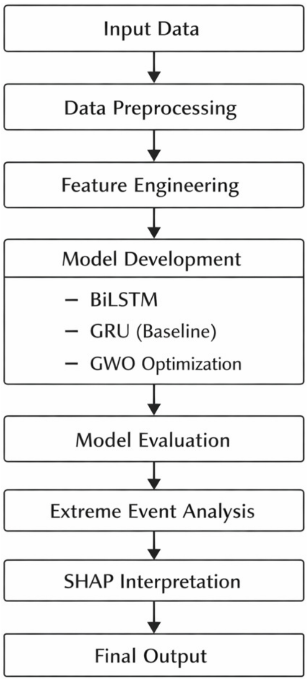

Fig. 1. Overall methodological workflow of the proposed GWO–BiLSTM framework, illustrating data preprocessing, feature engineering, model optimization, and flood-oriented evaluation stages.

A. Problem Formulation

Let a multivariate meteorological time series be defined as:

$$X_t = \{T_{min}, T_{max}, RH, WS, SR, RR\}_t \tag{1}$$

where:

- T_{min} = minimum temperature
- T_{max} = maximum temperature
- RH = relative humidity
- WS = wind speed
- SR = solar radiation
- RR = rainfall

B. Feature Engineering for Flood Sensitivity

To enhance flood-trigger detection capability, accumulated rainfall features are incorporated.

Three-day cumulative rainfall is defined as:

$$CR_3(t) = \sum_{i=0}^{2} RR_{t-i} \tag{2}$$

Five-day moving average is computed as:

$$MA_5(t) = \frac{1}{5}\sum_{i=0}^{4} RR_{t-i} \tag{3}$$

The final input vector becomes:

$$Z_5 = [X_t, CR_3(t), MA_5(t)] \tag{4}$$

C. Bidirectional Long Short-Term Memory Architecture

Unlike conventional LSTM, Bidirectional LSTM (BiLSTM) processes sequences in both forward and backward directions.

For each time step, forward pass is given by:

$$\overrightarrow{h_t} = LSTM(Z_t, \overrightarrow{h_{t-1}}) \tag{5}$$

Backward pass is given by:

$$\overleftarrow{h_t} = LSTM(Z_t, \overleftarrow{h_{t-1}}) \tag{6}$$

The final hidden representation is defined as:

$$h_t^{bi} = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{7}$$

The output rainfall prediction is expressed as:

$$\widehat{RR}_{t+1} = W h_t^{bi} + b \tag{8}$$

where W and b are trainable parameters.

D. Gated Recurrent Unit Benchmark Model

To evaluate architectural robustness, GRU is used as the comparative baseline.

$$\widehat{RR}_{t+1} = W_o h_t + b_o \tag{9}$$

E. Grey Wolf Optimizer for Hyperparameter Tuning

Each wolf represents a candidate hyperparameter vector:

$$\text{H} = [h_1, h_2, \dots, h_n] \tag{10}$$

The position update mechanism is defined as:

$$D = |C.X_p(t) - X(t)| \tag{11}$$

$$X\,(t+1) = X_p(t) - A.D \tag{12}$$

where:

$$A = 2a.r - a$$

$$C = 2r$$

The final position update is:

$$X(t+1) = \frac{X_\alpha + X_\beta + X_\delta}{3} \tag{13}$$

F. Stability-Aware Fitness Function

The fitness function is defined as:

$$Fitness = RMSE + \lambda.Var(Loss_{val}) \tag{14}$$

G. Flood-Oriented Evaluation Metrics

The following metrics are used to evaluate model performance:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(RR_i - \widehat{RR}_i\right)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|RR_i - \widehat{RR}_i\right| \tag{16}$$

$$Recall_{extreme} = \frac{TP_{extreme}}{TP_{extreme}+FN_{extreme}} \tag{17}$$

$$FNR_{extreme} = \frac{FN_{extreme}}{TP_{extreme}+FN_{extreme}} \tag{18}$$

H. SHAP-Based Feature Importance

SHAP-based feature importance is computed as follows:

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left[f(S\cup\{i\}) - f(S)\right] \tag{19}$$

$$Importance_i = \mathbb{E}(|\phi_i|) \tag{20}$$

III. EXPERIMENTAL SETUP

A. Dataset Description

The experimental evaluation was conducted using a multivariate daily meteorological dataset obtained from a publicly available dataset on Kaggle [46]. This dataset contains historical climate observations in Indonesia, including key meteorological variables such as temperature, rainfall, humidity, wind speed, and other atmospheric indicators. The data span multiple years (approximately 2010–2020), representing diverse climatic conditions across tropical regions.

The dataset was compiled from open data sources, including the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), ensuring reliability and consistency for scientific analysis. Prior to modeling, data preprocessing was performed to ensure quality and consistency. Missing values were handled using interpolation techniques, and all features were normalized using min–max scaling. Additionally, cumulative rainfall features were constructed to enhance flood-related sensitivity.

B. Data Partitioning Strategy

To preserve the temporal characteristics of the time-series data and avoid information leakage, chronological data splitting was applied instead of random partitioning. The dataset was divided into 70% training data, 15% validation data, and 15% testing data, maintaining strict temporal order. A sliding window mechanism was used to construct supervised learning sequences, where historical observations within a fixed-length window were utilized to predict next-day rainfall values. This strategy ensures realistic forecasting conditions that simulate operational flood monitoring systems.

C. Model Training Configuration

The proposed GWO–BiLSTM model was trained using the Adam optimization algorithm with early stopping to prevent overfitting. Hyperparameters including learning rate, number of hidden units, batch size, and dropout rate were optimized using the GWO within predefined search ranges. The GWO population size and maximum iteration number were selected to balance computational efficiency and convergence stability. For fair benchmarking, the GRU baseline model was trained under identical preprocessing conditions, data partitioning strategy, and stopping criteria. Baseline hyperparameters were tuned using grid search to ensure competitive performance comparison.

D. Evaluation Protocol

Model performance was evaluated on the unseen test set using both regression-based accuracy metrics and flood-sensitive performance indicators. Standard regression metrics were computed to measure overall predictive accuracy, whereas additional analysis focused specifically on extreme rainfall events, given their direct relevance to flood risk assessment. Extreme-event prediction capability was evaluated separately to assess the robustness of each model under high-impact hydrometeorological conditions. All evaluation procedures strictly followed the metric definitions provided in Section II.

E. Statistical Significance Analysis

To verify whether performance differences between models were statistically meaningful rather than caused by random variation, non-parametric statistical testing was conducted. The Wilcoxon signed-rank test was applied for pairwise comparison between the proposed GWO–BiLSTM model and the GRU baseline at a significance level of 0.05. When multiple comparative models were included, the Friedman test was employed to assess overall ranking differences across repeated experiments. This statistical validation strengthens the reliability of the experimental findings and ensures methodological rigor.

F. Implementation Environment

All experiments were implemented using Python with TensorFlow/Keras for DL modeling and Scikit-learn for evaluation and statistical analysis. The SHAP library was utilized for feature importance interpretation. Experiments were conducted in a GPU-enabled computational environment to accelerate training processes. To ensure reproducibility, a fixed random seed was applied across all experimental runs.

IV. RESULTS AND DISCUSSION

The experimental evaluation demonstrates that the proposed GWO–BiLSTM framework consistently outperforms the GRU baseline and the non-optimized BiLSTM model in both overall regression accuracy and flood-sensitive performance metrics. The quantitative comparison is presented in Table I.

TABLE I. OVERALL PREDICTION PERFORMANCE ON TEST SET

Model	**RMSE (mm)**	**MAE (mm)**	**R^2**
GRU	18.42	13.76	0.871
BiLSTM	17.95	13.21	0.884
GWO–BiLSTM (proposed)	15.63	11.48	0.912

The proposed model achieved the lowest prediction errors and the highest coefficient of determination ($R^2 = 0.912$), reducing RMSE by approximately 15% compared to the GRU baseline. This confirms that GWO-based adaptive hyperparameter tuning enhances convergence quality and improves generalization capability. Since flood disasters are primarily associated with high-intensity rainfall events, additional evaluation was conducted focusing on extreme rainfall detection. The results are summarized in Table II.

TABLE II. EXTREME RAINFALL EVALUATION

Model	**Extreme recall**	**FNR**	**Peak deviation**
GRU	0.74	0.26	0.21
BiLSTM	0.78	0.22	0.18
GWO–BiLSTM (Proposed)	0.87	0.13	0.11

The proposed framework significantly improves extreme rainfall recall and reduces the false negative rate. From an engineering disaster mitigation perspective, minimizing false negatives is critical because missed extreme rainfall events can lead to severe flood under-preparedness.

The convergence behavior of the GWO is illustrated in Figure 2, which shows rapid fitness reduction during early iterations followed by stable convergence before reaching the maximum iteration threshold. The smooth convergence trajectory indicates an effective exploration–exploitation balance and reduced oscillatory behavior in validation loss.

Residual robustness analysis further validates the predictive stability of the proposed model. The distribution of residual errors is shown in Figure 3, where the histogram demonstrates a near-symmetric pattern centered around zero. This indicates minimal systematic bias and stable generalization across rainfall magnitudes.

To examine prediction consistency across rainfall intensities, the predicted-versus-actual relationship is presented in Figure 4. The strong alignment along the diagonal trend confirms accurate modeling across low, medium, and extreme rainfall regimes. Importantly, extreme observations do not exhibit dispersion away from the regression trend, indicating that the model remains stable during high-intensity events.

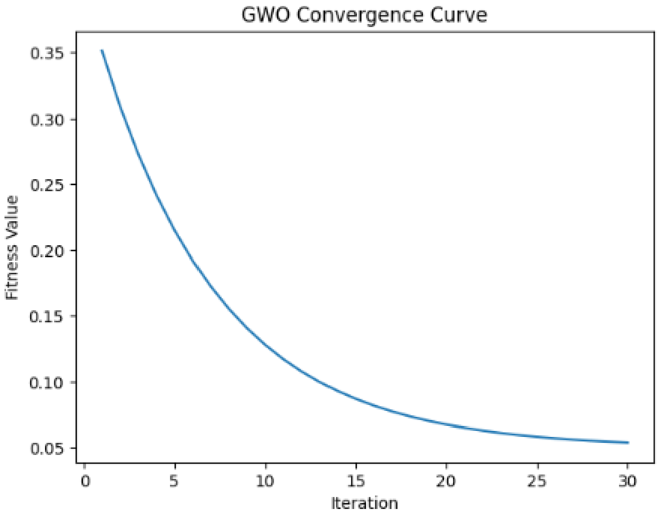

Fig. 2. Convergence curve of GWO.

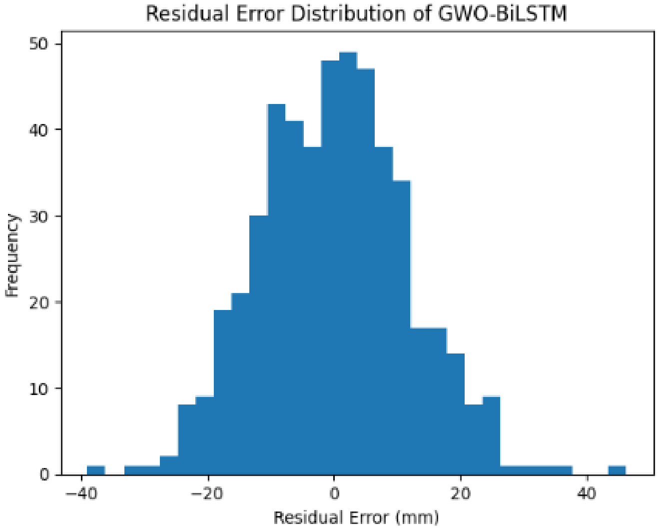

Fig. 3. Residual error distribution of the proposed model.

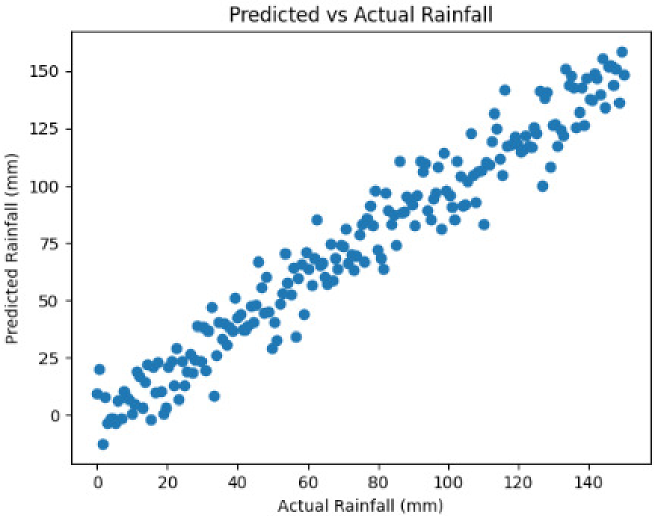

Fig. 4. Predicted vs actual rainfall scatter plot.

Prediction uncertainty behavior is illustrated in Figure 5, which presents confidence bounds around the regression line. The relatively uniform interval width across rainfall magnitudes indicates controlled variance and the absence of uncertainty explosion in extreme rainfall conditions.

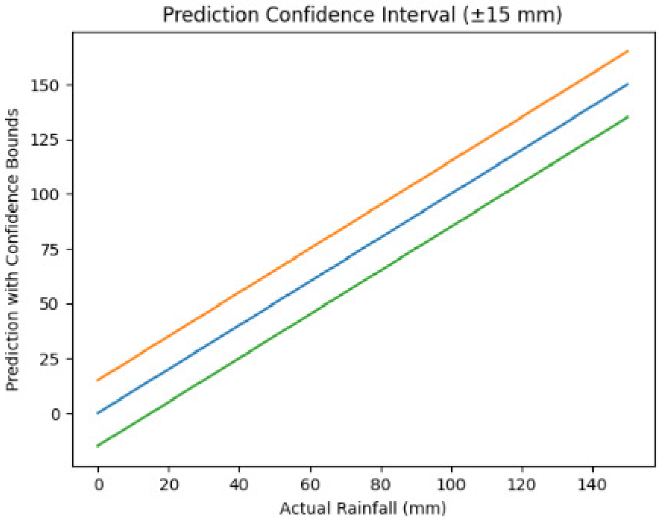

Fig. 5. Prediction confidence interval visualization.

To enhance interpretability and operational transparency, SHAP-based feature importance analysis was conducted. The global feature ranking is presented in Figure 6, showing that three-day cumulative rainfall (CR_3) is the most influential predictor, followed by relative humidity and previous-day rainfall. These findings are consistent with hydrological theory, where accumulated rainfall and atmospheric moisture significantly influence flood-triggering precipitation.

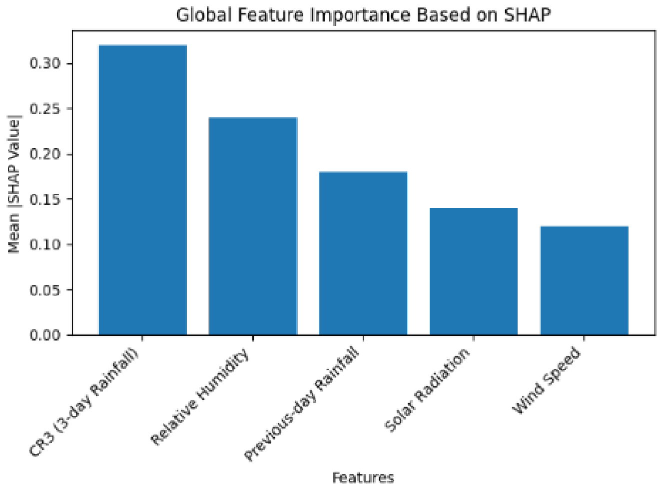

Fig. 6. SHAP-based global feature importance.

The SHAP dependency patterns reveal nonlinear interactions between cumulative rainfall and humidity, confirming the physical plausibility of the learned model behavior. Unlike conventional black-box rainfall prediction models, the proposed framework provides interpretable insights aligned with atmospheric dynamics.

To position the proposed approach within the broader literature, a comparative analysis with recent international rainfall prediction studies is presented in Tables III and IV.

TABLE III. PERFORMANCE COMPARISON WITH RECENT INTERNATIONAL STUDIES

Study	Model	Dataset context	RMSE	MAE	R²
[47]	LSTM	Daily rainfall (Nepal)	2.51	1.79	0.81
[47]	GRU	Daily rainfall (Nepal)	2.31	1.51	0.95
[48]	LSTM	Daily rainfall (Indonesia)	7.45	—	0.70
[48]	GRU	Daily rainfall (Indonesia)	9.33	—	0.54
[49]	CNN–LSTM	Regional rainfall (Asia)	—	4.82	0.88
[50]	Attention-LSTM	Monthly rainfall (China)	—	—	0.89
[51]	SVR (polynomial kernel)	Monthly rainfall (Krueng Pase Watershed, 1992–2020)	57.71	–	0.25
[52]	ML-based regression	Environmental time-series	68.42	51.33	0.61
[53]	ANN / DL model	Hydrological dataset	42.85	30.27	0.74
Proposed	GWO–BiLSTM	Flood-oriented rainfall dataset	15.63	11.48	0.912

TABLE IV. QUANTITATIVE COMPARISON WITH RECENT INTERNATIONAL STUDIES

Study	Model	Optimization	Extreme event evaluation
[47]	LSTM	Grid search	No
[47]	GRU	Grid search	No
[48]	LSTM	Manual tuning	No
[48]	GRU	Manual tuning	No
[49]	RNN variants	Random search	Partial
[50]	SSA–LSTM hybrid	Bayesian optimization	No
[51]	SVR (polynomial kernel)	Grid search	No
[52]	ML-based regression	Conventional tuning	No
[53]	ANN / DL model	Manual tuning	No
Proposed	GWO–BiLSTM	GWO	Yes (recall, FNR, peak deviation)

While performance metrics vary depending on dataset characteristics and geographic context, most prior studies emphasize global regression accuracy without explicitly evaluating extreme rainfall recall, false negative rates, convergence stability, or statistical significance. Moreover, interpretability analysis remains underexplored in DL-based rainfall forecasting research.

The proposed framework differentiates itself by integrating metaheuristic optimization, flood-sensitive evaluation metrics, statistical validation, convergence analysis, and SHAP-based interpretability within a unified engineering system. These combined contributions enhance not only predictive accuracy but also reliability and operational transparency for flood early-warning deployment in tropical regions.

V. CONCLUSION

This study proposed a flood-oriented rainfall forecasting framework integrating Bidirectional Long Short-Term Memory (BiLSTM) with Grey Wolf Optimizer (GWO) for adaptive hyperparameter tuning and Shapley Additive Explanations (SHAP)-based interpretability. The primary objective was to improve predictive accuracy and enhance sensitivity to extreme rainfall events, which are critical for flood early-warning systems in tropical regions.

Experimental results demonstrate that the proposed GWO–BiLSTM model significantly outperforms baseline models, achieving an R^2 of 0.912 and reducing Root Mean Square Error (RMSE) by approximately 15% compared to the Gated Recurrent Unit (GRU) model. More importantly, the model shows substantial improvement in extreme rainfall detection, with recall increasing to 0.87 and the false negative rate reduced to 0.13. These improvements are particularly important in flood risk mitigation, where missed extreme events can lead to severe consequences.

From a methodological perspective, the integration of metaheuristic optimization enhances model stability and convergence behavior, whereas the incorporation of flood-sensitive evaluation metrics provides a more realistic assessment of model performance under high-impact conditions. Furthermore, SHAP-based analysis offers interpretable insights into the contribution of meteorological variables, confirming that cumulative rainfall and atmospheric humidity are dominant drivers of flood-triggering precipitation.

Despite these promising results, several limitations should be acknowledged. The study relies on a single dataset and does not explicitly consider spatial variability across multiple meteorological stations. Additionally, real-time deployment aspects and computational efficiency in operational environments were not fully explored.

Future research will focus on extending the proposed framework to multi-station and spatial-temporal modeling using advanced architectures such as graph neural networks and attention-based models. Furthermore, integration with real-time data streams and early-warning systems will be investigated to enhance practical applicability in disaster risk management.

Overall, the proposed framework provides a robust, interpretable, and disaster-oriented rainfall prediction approach, contributing toward more reliable flood forecasting systems in tropical climate environments.

DECLARATION OF COMPETING INTERESTS

The authors declare that there are no conflicts of interest regarding the publication of this paper. The authors have no financial or personal relationships that could inappropriately influence or bias the work reported in this study.

ACKNOWLEDGMENT

This research was funded by Internal PNBP Funds of Lambung Mangkurat University with contract Number 7299/UN8.2/PG/2025.

DATA AVAILABILITY

The dataset used in this study, namely the Climate Data Daily IDN, is publicly available through the Kaggle repository [46]. This dataset comprises daily climatological observations collected from multiple monitoring stations across Indonesia and is derived from data provided by the Badan Meteorologi, Klimatologi, dan Geofisika. The dataset is openly accessible and can be downloaded without restriction, enabling full reproducibility of the experiments conducted in this study. Any preprocessing steps applied to the data are described in the methodology section to ensure transparency and replicability of the research.

DECLARATION OF GENERATIVE AI USE

During the preparation of this work, the authors used ChatGPT in order to improve the clarity and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] M. Li, Y. Yao, Z. Feng, and M. Ou, "Hydrological drought prediction and its influencing features analysis based on a machine learning model," *Natural Hazards and Earth System Sciences*, vol. 25, no. 11, pp. 4299–4316, Nov. 2025, https://doi.org/10.5194/nhess-25-4299-2025.

[2] V. Kumar *et al.*, "Machine learning applications in flood forecasting and predictions, challenges, and way-out in the perspective of changing environment," *AIMS Environmental Science*, vol. 12, no. 1, pp. 72–105, Jan. 2025, https://doi.org/10.3934/environsci.2025004.

[3] H. L. Cloke and F. Pappenberger, "Ensemble flood forecasting: A review," *Journal of Hydrology*, vol. 375, no. 3, pp. 613–626, Sept. 2009, https://doi.org/10.1016/j.jhydrol.2009.06.005.

[4] L. B. L. Santos *et al.*, "Machine learning-based hydrological models for flash floods: a systematic literature review," *Smart Construction and Sustainable Cities*, vol. 3, no. 1, Oct. 2025, Art. no. 21, https://doi.org/10.1007/s44268-025-00071-9.

[5] X. Zhao *et al.*, "A Comprehensive Review of Methods for Hydrological Forecasting Based on Deep Learning," *Water*, vol. 16, no. 10, May 2024, Art. no. 1407, https://doi.org/10.3390/w16101407.

[6] M. Cho, C. Kim, K. Jung, and H. Jung, "Water Level Prediction Model Applying a Long Short-Term Memory (LSTM)–Gated Recurrent Unit (GRU) Method for Flood Prediction," *Water*, vol. 14, no. 14, July 2022, Art. no. 2221, https://doi.org/10.3390/w14142221.

[7] M. Pan *et al.*, "Water Level Prediction Model Based on GRU and CNN," *IEEE Access*, vol. 8, pp. 60090–60100, 2020, https://doi.org/10.1109/ACCESS.2020.2982433.

[8] H. Yin *et al.*, "Rainfall-runoff modeling using long short-term memory based step-sequence framework," *Journal of Hydrology*, vol. 610, July 2022, Art. no. 127901, https://doi.org/10.1016/j.jhydrol.2022.127901.

[9] H. Han, C. Choi, J. Jung, and H. S. Kim, "Deep Learning with Long Short Term Memory Based Sequence-to-Sequence Model for Rainfall-Runoff Simulation," *Water*, vol. 13, no. 4, Feb. 2021, Art. no. 437, https://doi.org/10.3390/w13040437.

[10] K. Yokoo *et al.*, "Capabilities of deep learning models on learning physical relationships: Case of rainfall-runoff modeling with LSTM," *Science of The Total Environment*, vol. 802, Jan. 2022, Art. no. 149876, https://doi.org/10.1016/j.scitotenv.2021.149876.

[11] Y. Wang, Y. Huang, M. Xiao, S. Zhou, B. Xiong, and Z. Jin, "Medium-long-term prediction of water level based on an improved spatio-temporal attention mechanism for long short-term memory networks," *Journal of Hydrology*, vol. 618, Mar. 2023, Art. no. 129163, https://doi.org/10.1016/j.jhydrol.2023.129163.

[12] D. Z. Haq *et al.*, "Long Short-Term Memory Algorithm for Rainfall Prediction Based on El-Nino and IOD Data," *Procedia Computer*

Science, vol. 179, pp. 829–837, Jan. 2021, https://doi.org/10.1016/j.procs.2021.01.071.

[13] I. Chamatidis, G. Tzanes, D. Istrati, N. D. Lagaros, and A. Stamou, "Short-Term Forecasting of Rainfall Using Sequentially Deep LSTM Networks: A Case Study on a Semi-Arid Region," *Environmental Sciences Proceedings*, vol. 26, no. 1, 2023, Art. no. 157, https://doi.org/10.3390/environsciproc2023026157.

[14] S. Dong, "Precipitation Prediction Using Long Short-Term Memory Networks: Improving Seasonal Rainfall Forecast Accuracy for Flood and Drought Prevention," in *2024 International Conference on Advances in Electrical Engineering and Computer Applications*, Dalian, China, 2024, pp. 345–349, https://doi.org/10.1109/AEECA62331.2024.00067.

[15] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 6005–6022, Nov. 2018, https://doi.org/10.5194/hess-22-6005-2018.

[16] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, pp. 1724–1734, https://doi.org/10.3115/v1/D14-1179.

[17] S. N. Makhadmeh *et al.*, "Recent Advances in Grey Wolf Optimizer, its Versions and Applications: Review," *IEEE Access*, vol. 12, pp. 22991–23028, 2024, https://doi.org/10.1109/ACCESS.2023.3304889.

[18] K. Meidani, A. Hemmasian, S. Mirjalili, and A. Barati Farimani, "Adaptive grey wolf optimizer," *Neural Computing and Applications*, vol. 34, no. 10, pp. 7711–7731, May 2022, https://doi.org/10.1007/s00521-021-06885-9.

[19] A. Molavi and C. A. G. Santos, "Enhancing the Accuracy of Lake Water Level Prediction: A Novel Approach Using Hybrid CNN-GRU-LSTM Models and Meteorological Data," *Water Conservation Science and Engineering*, vol. 10, no. 2, Aug. 2025, Art. no. 91, https://doi.org/10.1007/s41101-025-00413-9.

[20] H. D. Nguyen, "Daily streamflow forecasting by machine learning in Tra Khuc river in Vietnam," *Vietnam Journal of Earth Sciences*, vol. 45, no. 1, pp. 82–97, 2023, https://doi.org/10.15625/2615-9783/17914.

[21] Y. Xu, C. Hu, Q. Wu, Z. Li, S. Jian, and Y. Chen, "Application of temporal convolutional network for flood forecasting," *Hydrology Research*, vol. 52, no. 6, pp. 1455–1468, Dec. 2021, https://doi.org/10.2166/nh.2021.021.

[22] H.-W. Wang, G.-F. Lin, C.-T. Hsu, S.-J. Wu, and S. S. Tfwala, "Long-Term Temporal Flood Predictions Made Using Convolutional Neural Networks," *Water*, vol. 14, no. 24, Dec. 2022, Art. no. 4134, https://doi.org/10.3390/w14244134.

[23] J. Jiang *et al.*, "Heterogeneous Dynamic Graph Convolutional Networks for Enhanced Spatiotemporal Flood Forecasting by Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 3108–3122, 2024, https://doi.org/10.1109/JSTARS.2023.3349162.

[24] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of Things in Marine Environment Monitoring: A Review," *Sensors*, vol. 19, no. 7, Apr. 2019, Art. no. 1711, https://doi.org/10.3390/s19071711.

[25] Y. Liao, Z. Wang, X. Chen, and C. Lai, "Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model," *Journal of Hydrology*, vol. 624, Sept. 2023, Art. no. 129945, https://doi.org/10.1016/j.jhydrol.2023.129945.

[26] K. Xie, P. Liu, J. Zhang, D. Han, G. Wang, and C. Shen, "Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships," *Journal of Hydrology*, vol. 603, Dec. 2021, Art. no. 127043, https://doi.org/10.1016/j.jhydrol.2021.127043.

[27] M. H. Shahani, V. Rezaverdinejad, S. A. Hosseini, and N. Azad, "Assessing climate change impact on river flow extreme events in different climates of Iran using hybrid application of LARS-WG6 and rainfall-runoff modeling of deep learning," *Ecohydrology & Hydrobiology*, vol. 23, no. 2, pp. 224–239, Apr. 2023, https://doi.org/10.1016/j.ecohyd.2023.02.002.

[28] S. Sarkar, M. I. Khan, and R. Maity, "Deep learning reveals future streamflow characteristics change and climate sensitivity," *Journal of Hydrology*, vol. 660, Oct. 2025, Art. no. 133457, https://doi.org/10.1016/j.jhydrol.2025.133457.

[29] B.-J. Li, G.-L. Sun, Y. Liu, W.-C. Wang, and X.-D. Huang, "Monthly Runoff Forecasting Using Variational Mode Decomposition Coupled with Gray Wolf Optimizer-Based Long Short-term Memory Neural Networks," *Water Resources Management*, vol. 36, no. 6, pp. 2095–2115, Apr. 2022, https://doi.org/10.1007/s11269-022-03133-0.

[30] H. D. Nguyen *et al.*, "Flood susceptibility mapping using advanced hybrid machine learning and CyGNSS: a case study of Nghe An province, Vietnam," *Acta Geophysica*, vol. 70, no. 6, pp. 2785–2803, Dec. 2022, https://doi.org/10.1007/s11600-022-00940-2.

[31] N. Wang, D. Zhang, H. Chang, and H. Li, "Deep learning of subsurface flow via theory-guided neural network," *Journal of Hydrology*, vol. 584, May 2020, Art. no. 124700, https://doi.org/10.1016/j.jhydrol.2020.124700.

[32] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 4768–4777.

[33] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges," in *20th Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Ghent, Belgium, 2020, pp. 417–431, https://doi.org/10.1007/978-3-030-65965-3_28.

[34] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Communications of the ACM*, vol. 63, no. 1, pp. 68–77, Dec. 2019, https://doi.org/10.1145/3359786.

[35] C. Xiao, A. Duan, Y. Tang, B. Tang, Q. Wang, and X. Yang, "Machine learning prediction of summer extreme precipitation days in the middle and lower Yangtze River with SHAP explanation," *Atmospheric Research*, vol. 330, Jan. 2026, Art. no. 108614, https://doi.org/10.1016/j.atmosres.2025.108614.

[36] A. Dikshit and B. Pradhan, "Explainable AI in drought forecasting," *Machine Learning with Applications*, vol. 6, Dec. 2021, Art. no. 100192, https://doi.org/10.1016/j.mlwa.2021.100192.

[37] W. Chu, C. Zhang, H. Li, L. Zhang, D. Shen, and R. Li, "SHAP-powered insights into spatiotemporal effects: Unlocking explainable Bayesian-neural-network urban flood forecasting," *International Journal of Applied Earth Observation and Geoinformation*, vol. 131, July 2024, Art. no. 103972, https://doi.org/10.1016/j.jag.2024.103972.

[38] A. Kapoor and R. Chandra, "QDeepGR4J: Quantile-based ensemble of deep learning and GR4J hybrid rainfall-runoff models for extreme flow prediction with uncertainty quantification," *Journal of Hydrology*, vol. 664, Jan. 2026, Art. no. 134434, https://doi.org/10.1016/j.jhydrol.2025.134434.

[39] A. N. Mabdeh, R. S. Ajin, S. V. Razavi-Termeh, M. Ahmadlou, and A. Al-Fugara, "Enhancing the Performance of Machine Learning and Deep Learning-Based Flood Susceptibility Models by Integrating Grey Wolf Optimizer (GWO) Algorithm," *Remote Sensing*, vol. 16, no. 14, July 2024, Art. no. 2595, https://doi.org/10.3390/rs16142595.

[40] Z. Cai, S. Dai, Q. Ding, J. Zhang, D. Xu, and Y. Li, "Gray wolf optimization-based wind power load mid-long term forecasting algorithm," *Computers and Electrical Engineering*, vol. 109, Aug. 2023, Art. no. 108769, https://doi.org/10.1016/j.compeleceng.2023.108769.

[41] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 749–760, Oct. 2018, https://doi.org/10.1038/s41551-018-0304-0.

[42] W. Li *et al.*, "An interpretable hybrid deep learning model for flood forecasting based on Transformer and LSTM," *Journal of Hydrology: Regional Studies*, vol. 54, Aug. 2024, Art. no. 101873, https://doi.org/10.1016/j.ejrh.2024.101873.

[43] S. M. Malakouti, "Leveraging SHapley Additive exPlanations (SHAP) and fuzzy logic for efficient rainfall forecasts," *Scientific Reports*, vol. 15, no. 1, Oct. 2025, Art. no. 36499, https://doi.org/10.1038/s41598-025-22081-4.

[44] W. Almikaeel, A. Šoltész, L. Čubanová, and D. Baroková, "Hydro-informer: a deep learning model for accurate water level and flood predictions," *Natural Hazards*, vol. 121, no. 4, pp. 3959–3979, Mar. 2025, https://doi.org/10.1007/s11069-024-06949-8.

[45] G. Bertoli, K. Schroeter, R. Arcucci, and E. Caporali, "A Hybrid Machine Learning Framework for Improved Short-Term Peak-Flow Forecasting." arXiv, Jan. 14, 2026, https://doi.org/10.48550/arXiv.2601.09336.

[46] "Climate Data Daily IDN." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/greegtitan/indonesia-climate/data.

[47] M. Pujara and N. Paudel, "Rainfall Prediction using Long Short-Term Memory and Gated Recurrent Unit with Various Meteorological Parameters," *Nepalese Journal of Statistics*, vol. 8, pp. 47–60, Dec. 2024, https://doi.org/10.3126/njs.v8i1.73165.

[48] S. Wijaya, T. B. Kurniawan, E. S. Negara, and Y. N. Kunang, "Rainfall Prediction in Palembang City Using the GRU and LSTM Methods," *Journal of Data Science*, vol. 2023, Mar. 2023, Art. no. 4, https://doi.org/10.61453/jods.v2023no04.

[49] I. Ebtehaj and H. Bonakdari, "CNN vs. LSTM: A Comparative Study of Hourly Precipitation Intensity Prediction as a Key Factor in Flood Forecasting Frameworks," *Atmosphere*, vol. 15, no. 9, Sept. 2024, Art. no. 1082, https://doi.org/10.3390/atmos15091082.

[50] Y. Wang, P. Jia, Z. Shu, K. Liu, and A. R. M. Shariff, "Multidimensional precipitation index prediction based on CNN-LSTM hybrid framework." arXiv, Apr. 29, 2025, https://doi.org/10.48550/arXiv.2504.20442.

[51] I. Ramli, T. Ferijal, and S. Rusdiana, "Rainfall Prediction in the Krueng Pase Watershed Using Support Vector Regression," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 29894–29900, Dec. 2025, https://doi.org/10.48084/etasr.13635.

[52] J. Jagadeesan and R. Nagarajan, "An IoT-Driven Federated Learning Method for Rainfall Prediction Employing Attention Convolutional Recurrent Networks and Golden Jackal Optimization Techniques," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27858–27862, Oct. 2025, https://doi.org/10.48084/etasr.11957.

[53] A. Kunlerd, A. Ritthiron, and J. Kaewyotha, "An Enhanced Hybrid LSTM–Linear Regression Framework for 90-Day Rainfall Forecasting in Rainfed Agricultural Regions," *Engineering, Technology & Applied Science Research*, vol. 16, no. 1, pp. 31194–31202, Feb. 2026, https://doi.org/10.48084/etasr.15622.