Numerical Methods in Civil Engineering, 10-4 (2026) 1-11

Numerical Methods in Civil Engineering

Journal Homepage: https://nmce.kntu.ac.ir/

A Spatiotemporal CNN–LSTM Framework for Daily Precipitation Bias Correction Using Satellite and Reanalysis Data

Seyed Mohammad Miri[*], Mohammad Reza Kavianpour[**]

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
January 2026
Revised:
May 2026
Accepted:
May 2026

Keywords:
Satellite precipitation,
Rainfall estimation,
CNN-LSTM,
ERA5 reanalysis,
Spatiotemporal Deep Learning

Abstract:

Accurate daily precipitation estimates are critical for hydrological and civil engineering applications, especially in regions affected by short-duration rainfall events. Satellite-based products provide wide coverage but often show systematic biases compared with ground observations. This study presents a spatiotemporal convolutional neural network–long short-term memory (CNN–LSTM) framework for daily precipitation bias correction in Golestan Province, northern Iran, using satellite and reanalysis data. The proposed framework integrates PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) satellite precipitation with ERA5 reanalysis precipitation and 2-m air temperature to improve station-scale rainfall estimation at several synoptic stations. In order to represent the spatial organization of precipitation systems vividly, gridded precipitation from neighboring upstream cells was incorporated into the model structure. This configuration accounts for the dominant west-to-east movement of rainfall systems across the region, where precipitation commonly arrives with temporal delay and reduced intensity. In addition, temporal variations in near-surface air temperature were included to provide supplementary physical information associated with rainfall development. Model performance is evaluated using root mean square error (RMSE), correlation coefficient (CC), and mean error (ME). Categorical skill is assessed using probability of detection (POD) and false alarm ratio (FAR). The results indicate that the CNN–LSTM model consistently outperformed both conventional bias-correction approaches and baseline deep learning models. Incorporating spatial information from adjacent grid cells reduces errors and improves temporal consistency. Across the evaluated stations, the proposed model achieved the lowest RMSE values, with error reductions reaching 23% relative to the DNN benchmark. The framework also demonstrated improved capability in capturing the spatiotemporal evolution of precipitation events. Overall, the proposed approach provides a transferable and operationally practical framework for enhancing daily precipitation estimates in data-scarce regions. The improved precipitation estimates can support more reliable hydrological simulations, flood assessment, and water resources planning.

1. Introduction

Accurate daily precipitation information is fundamental to hydrological modeling, urban drainage design, flood risk assessment, and water resources management, particularly in regions affected by short-duration and high-intensity rainfall events. In arid and semi-arid environments such as Golestan Province in northern Iran, the scarcity of ground-based observations limits the reliability of rainfall estimates. Satellite-derived precipitation products, including PERSIANN, provide broad spatial coverage and near-real-time monitoring capability; however, they commonly

* *Corresponding author: Ph.D. Candidate, Faculty of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email: m.miri@email.kntu.ac.ir*

** *Professor, Faculty of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email: kavianpour@kntu.ac.ir*

https://doi.org/10.61882/NMCE.2601.1124

exhibit systematic discrepancies relative to in-situ observations, especially during convective storms and extreme rainfall conditions [1]. Although reanalysis datasets such as ERA5 offers temporally consistent and physically coherent precipitation fields, they may not capture local variability sufficiently [2,3]. These limitations underscore the need for robust bias-correction and precipitation estimation approaches.

Numerous studies have assessed the performance of satellite-based precipitation products such as PERSIANN, as well as reanalysis datasets including ERA5, through comparisons with ground-based observations and gauge-derived precipitation records. For example, in the Tambo Basin of Peru, PERSIANN-CDR (Climate Data Record) exhibited relatively high Pearson correlation coefficients (CC ≈ 0.84–0.94) with observed monthly precipitation. Nevertheless, PERSIANN-CDR showed considerable bias (PBIAS ≈ 6.9–83.1%) and RMSE values ranging from approximately 22 to 39 mm/month, whereas ERA5, despite its relatively strong correlation with observations, generally overestimated total precipitation. [4]. In the Bandung Raya region of Indonesia, ERA5 generally outperformed PERSIANN-CDR in daily scale comparisons across RMSE, bias, and CC metrics, although both products displayed notable deviations from in-situ measurements [5]. Studies over semi-arid regions in southern India reported that at the grid scale, ERA5-Land often achieved superior RMSE, mean absolute error (MAE), and index of agreement compared with PERSIANN, though PERSIANN sometimes showed lower bias depending on the location [6]. Evaluations conducted at several stations in Ghana and Zambia showed that both reanalysis and satellite-based precipitation products are generally able to detect dry days with relatively high probability of detection ($POD > 0.7$). However, their ability to represent high-intensity rainfall events remains limited, indicating the need for bias correction and careful interpretation when these datasets are used for analyses involving extreme precipitation [7]. Additionally, global assessments including ERA5 and PERSIANN-CDR in Ethiopia reported RMSE values generally below about 25 mm and spatial patterns broadly consistent with observed climatology. However, discrepancies increased with higher precipitation intensity, suggesting that while both datasets can capture large-scale rainfall patterns, local calibration is often necessary for accurate quantitative estimates [8]. These findings have suggested that ERA5 and PERSIANN products provide useful precipitation information across a range of climatic conditions. However, their performance at finer temporal scales can vary considerably depending on regional characteristics and rainfall type, underscoring the need for bias correction and the integration of multiple data sources in h`ydrological applications. Traditional bias correction and downscaling methods, such as quantile mapping, linear scaling, and physical statistical models, are widely used to adjust satellite or reanalysis precipitation products. However, these methods often lack the ability to fully capture complex spatiotemporal dependencies and the nonlinear behavior of precipitation processes [2,9]. In particular, they typically ignore the spatiotemporal propagation of rainfall systems and physically meaningful auxiliary predictors, such as temperature gradients, which can indicate atmospheric dynamics and influence rainfall intensity.

In recent years, substantial progress has been made in the application of machine learning (ML) and deep learning (DL) techniques for precipitation estimation and bias correction. ML and DL approaches can model nonlinear relationships and integrate multiple data sources without strict assumptions. For example, CNN and LSTM architectures have been used to enhance the satellite precipitation estimates, capture spatial rainfall patterns, and improve detection of extreme events [10–17]. Studies have applied deep neural networks (DNNs) for bias correction of numerical weather prediction precipitation [11,14], Bi-LSTM frameworks for satellite precipitation [12,15], and multimodal learning combining radar, satellite, and gauge data [13,16-19]. Overall, these studies have shown that ML and DL approaches can enhance rainfall estimation and bias correction by learning complex spatial and temporal dependencies within hydroclimatic data. Among these approaches, hybrid architectures such as CNN–LSTM have gained increasing attention in hydrological research because they are able to represent both spatial structures and temporal dynamics in rainfall and runoff processes (e.g., Li et al. [20]). In many cases, integrating CNN-based spatial feature extraction with the sequence-learning capability of LSTM has produced more accurate rainfall predictions than using either architecture independently (e.g., Ahmed et al. [21]).

Despite these advances, relatively few studies have examined sub-daily precipitation bias correction in data-scarce regions such as Iran. In particular, applications that employ deep spatiotemporal models while jointly integrating satellite and reanalysis precipitation data remain limited. Moreover, the potential impact of spatial neighborhood information and temperature gradients in improving precipitation estimates remains underexplored. In Golestan Province, rainfall systems typically move from west to east, often with a noticeable time lag and changes in intensity. This behavior suggests that including information from upstream grid cells may help improve the accuracy of precipitation modeling.

This study proposes a spatiotemporal CNN–LSTM model that integrates PERSIANN and ERA5 precipitation and temperature data to estimate ground-based precipitation at synoptic stations. The model accounts for west–east propagation patterns, spatial information from neighboring grids, and temperature gradients as auxiliary predictors. The framework is applied to daily rainfall estimation, and its performance is evaluated using multiple statistical and categorical metrics to demonstrate its effectiveness.

2. Study Area and Data

2.1 Study Area

Located in northeastern Iran, Golestan Province is characterized by substantial climatic and topographic variability. The province stretches from the southern coast of the Caspian Sea toward the northern foothills of the Alborz Mountains, creating marked spatial differences in precipitation distribution. Its landscape ranges from low-lying coastal plains in the north to mountainous and hilly areas in the south, and these elevation contrasts play an important role in shaping regional atmospheric circulation and rainfall patterns. The climate of Golestan Province varies from humid and semi-humid conditions in the northern and western regions to semi-arid conditions in the southern and eastern parts. The proximity to the Caspian Sea, complex topography, and interaction between maritime and continental air masses contribute to frequent short-duration and high-intensity rainfall events, making the region particularly suitable for short-term rainfall forecasting studies.

In this study, rainfall observations from eight synoptic meteorological stations across Golestan Province are used (Figure 1). The spatial distribution of the stations allows for adequate representation of the province's diverse climatic and topographic conditions. High-resolution rainfall data were obtained at a 10-minute temporal interval for the period 2019–2025. To investigate rainfall variability, the original data were aggregated into daily rainfall time series.

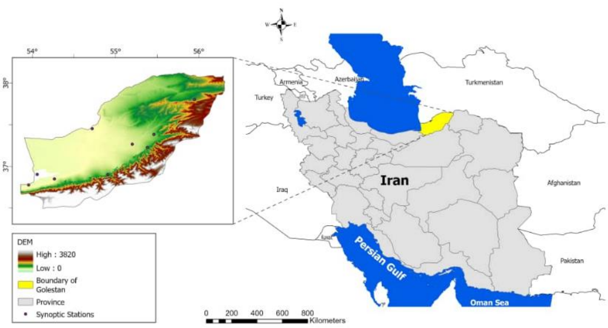

Fig. 1: Location of Golestan Province in northeastern Iran and spatial distribution of the eight synoptic meteorological stations used in this study

2.2 Datasets

Ground-based observations from synoptic meteorological stations were used as the reference dataset for evaluating rainfall estimates in this study. Gridded precipitation products from PERSIANN-UNet and the ERA5 reanalysis were employed as input predictors, enabling an assessment of how accurately these datasets reproduce the observed rainfall patterns across the Golestan Province.

Gridded precipitation estimates were obtained from PERSIANN-UNet dataset, a deep learning–based rainfall product within the PERSIANN family of satellite precipitation datasets. PERSIANN-UNet is designed to improve the spatial representation of precipitation, particularly over regions with complex terrain. For this study, rainfall data covering the Golestan Province were extracted and temporally aggregated to daily scales to ensure consistency with the aggregated synoptic observations.

We used ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides hourly gridded total precipitation and 2-m air temperature (t2m) for the study period. ERA5 data were extracted for the study area and aggregated into daily resolutions. These datasets were used to estimate rainfall at the synoptic stations. They were also used to assess ERA5's performance against observed measurements.

In addition to absolute temperature values, the temperature gradient was calculated as the temporal difference between consecutive aggregated time steps:

$$GT_t = T_{2m}(t) - T_{2m}(t-1) \quad (1)$$

The temperature tendency was included as an additional explanatory variable to examine whether short-term thermal variations help improve rainfall estimates.

Table 1: Statistical analysis of daily rainfall data

Location	Mean	Median	Std	Max
Aliabad	1.75	0.00	5.60	79.80
Bandargaz	1.38	0.00	5.26	89.10
Bandartorkaman	1.15	0.00	5.25	129.2
Gonbad	1.06	0.00	4.15	98.10
Hashemabad	1.22	0.00	5.08	117.7
Inchebroon	0.60	0.00	2.75	55.10
MaravehTappeh	0.75	0.00	3.30	57.80
Minoodasht	2.06	0.00	7.13	106.3

Table 1 summarizes the statistical characteristics of daily rainfall at the eight stations. Mean daily rainfall values are generally low, ranging from 0.60 mm at Inchebroon to 2.06 mm at Minoodasht, reflecting the region's overall dry conditions. Median values are zero at all stations, indicating that dry days dominate the record. Standard deviations vary more widely, with Minoodasht showing the highest variability (7.13 mm) and Inchebroon the lowest (2.75 mm), suggesting that rainfall events are more sporadic in some locations. Maximum daily rainfall highlights extreme events, with Bandartorkaman reaching 129.2 mm and Aliabad 79.8 mm. Overall, these numbers show strong heterogeneity in rainfall intensity and variability across stations, with some sites experiencing occasional heavy rainfall while most days remain dry.

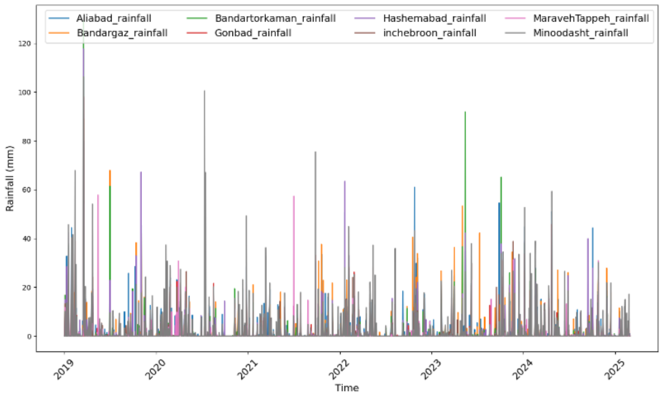

Fig. 2: Time series of daily rainfall records at different synoptic stations

Figure 2 shows the time series of observed daily rainfall across the eight synoptic stations in the Golestan Province. The data reveal substantial temporal variability and frequent short-duration events during the 2019–2025 period. Rainfall magnitudes are generally low, with intermittent peaks, reflecting the semi-arid nature of much of the region. Stations located in more humid or topographically favorable areas occasionally record higher-intensity events. Clear seasonal and interannual variability is also evident, with certain periods, such as 2023 and 2024, showing more pronounced rainfall activity at multiple stations. Overall, as illustrated in Figure 2, the dataset reflects the considerable variability of rainfall among stations. They also emphasize the need for high-resolution temporal observations to accurately capture rainfall patterns in the Golestan Province.

3. Methodology

3.1 Data Preprocessing

Three primary data sources were utilized in this study: PERSIANN-UNet satellite precipitation estimates, ERA5 reanalysis data, and ground-based observational records from eight meteorological stations across the Golestan Province.

PERSIANN-Unet was selected among PERSIANN precipitation products due to its strong performance in recent studies and its effectiveness in representing sub-daily spatial precipitation patterns. The PERSIANN-Unet product, with temporal resolution of 6-hourly and spatial resolution of 0.25°×0.25°, was obtained for the period 2019-2025. Consistent with common practices in hydrological studies using satellite-derived precipitation, the data were aggregated to a 24-hour time step. Simple summation was used to preserve total precipitation accumulation.

ERA5 hourly reanalysis data were obtained at a spatial resolution of 0.25° × 0.25°, including total precipitation (tp) and 2-meter air temperature (t2m) variables for the same temporal coverage. After aggregating to a 24-hour interval, both precipitation and temperature fields were extracted.

Observational precipitation data were collected from eight synoptic stations located across the Golestan Province. These records were aggregated to 24-hourly accumulation to ensure temporal consistency with satellite and reanalysis inputs. Following temporal splitting recommendations, the study period was divided into training (70%), validation (15%), and test (15%) subsets. This approach maintains chronological integrity while allowing full calibration of the dataset.

All datasets were quality-controlled to remove missing values and unrealistic precipitation intensities. Spatial coordinates (latitude, longitude, elevation) for each station were recorded to facilitate the extraction of corresponding grid cells from PERSIANN-UNet and ERA5 datasets.

3.2 Spatiotemporal Feature Construction

One important feature of the proposed framework is the explicit representation of spatiotemporal propagation dynamics. In particular, the model incorporates information from neighboring grid cells together with temporal lag features, both of which can contribute to improved precipitation estimation accuracy in deep learning applications.

3.3 Deep Learning Models

To assess the relative contribution of spatial and temporal feature extraction, three deep learning architectures were implemented and systematically compared. These comprised a fully connected deep neural network (DNN), a long short-term memory (LSTM) network designed to capture temporal dependencies, and a hybrid convolutional neural network–long short-term memory (CNN–LSTM) model intended to jointly represent spatial structure and temporal evolution in precipitation fields.

3.3.1 Deep Neural Network (DNN)

The DNN was implemented as a baseline model that does not explicitly account for either spatial structure or temporal dependencies, instead treating all input predictors as a flattened feature vector. The network consisted of four fully connected layers with 128, 64, 32, and 16 neurons, respectively. Each layer was followed by a ReLU activation function and a dropout layer (rate = 0.4) to reduce overfitting and improve generalization. The output layer was a single neuron with linear activation for precipitation regression. This model provides a benchmark for assessing the added value of spatiotemporal modeling.

3.3.2 Long Short-Term Memory (LSTM)

The LSTM model was designed to capture temporal dependencies while treating spatial features as independent inputs. The architecture included two LSTM layers with 64 and 32 hidden units, respectively, followed by a dense layer (16 neurons) and a final output neuron. Temporal sequences were formed by stacking lagged features along the time dimension. This model tests whether temporal memory alone, without explicit spatial convolution, can improve precipitation estimation.

3.3.3 CNN-LSTM Hybrid Model

The CNN–LSTM model forms the central part of the proposed framework, combining convolutional layers to extract spatial rainfall patterns with recurrent processing to represent temporal variability and sequence dependence. This architecture has been successfully applied for sub-daily precipitation estimation and bias correction in recent studies [22-24]. CNN extracts spatial patterns and local relationships from gridded atmospheric variables related to precipitation [25]. LSTM captures temporal dependencies and memory effects in hydro-climatic processes, allowing the model to account for delayed impacts of previous atmospheric conditions [26]. By combining CNN and LSTM, the model can capture both spatial patterns and temporal dependencies, providing a more effective representation of the processes that influence precipitation variability.

The model architecture consists of:

1) **Convolutional Layers:** Two 2D convolutional layers with 32 and 64 filters, kernel size 3×3, and ReLU activation were applied to the 3×3 spatial input grids. These layers were designed to capture local spatial structures in the precipitation field, including gradients, neighborhood correlations, and short-range propagation patterns. Because the spatial window is relatively small, max-pooling operations were not applied in order to avoid losing potentially important spatial information.

2) **LSTM Layers:** The spatial feature maps were subsequently flattened and provided as input to two LSTM layers with 64 and 32 hidden units, respectively, allowing the model to represent the temporal evolution of precipitation signals. Bidirectional LSTM configurations were also examined during preliminary testing; however, they were not retained in the final architecture because the improvement in predictive skill was limited compared with the additional computational cost.

3) **Dense Layers:** Two fully connected layers with 32 and 16 neurons, followed by dropout (0.3), were added for final nonlinear transformation.

4) **Output Layer:** A single neuron with linear activation produced the bias-corrected precipitation estimate for the target station.

The input shape for the CNN-LSTM model was [batch_size, time_steps, height, width, features]. Here, height and width correspond to the 3×3 spatial window, and features include PERSIANN, ERA5 precipitation, temperature, and gradients. This design choice is consistent with several recent studies in hydro-climatic modeling and downscaling that employ small convolution kernels to efficiently learn spatial features [25]. This design allows the model to learn spatiotemporal dependencies critical for accurate precipitation estimation.

3.4 Model Training and Evaluation

3.4.1 Training Configuration

All models were implemented in TensorFlow/Keras and trained on NVIDIA GPU with 16 GB RAM. The Mean Squared Error (MSE) loss function was adopted as the primary objective:

$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

where y_i is the observed precipitation and $\hat{y}_i$ is the model prediction.

To address class imbalance between dry and rainy days, a weighted loss function was explored following recent recommendations [27]:

$$\text{Weighted Loss} = \frac{1}{N}\sum_{i=1}^{N}w_i(y_i - \hat{y}_i)^2 \tag{3}$$

where $w_i = 2$ if $y_i > 0.1$ mm (rainy), otherwise $w_i = 1$ (dry). This approach emphasizes minority rainy events without excessive penalization of dry forecasts.

The models were trained using a batch size of 256 and the Adam optimizer with a learning rate of 0.0001. A maximum of 200 training epochs was allowed; however, early stopping based on validation loss, with a patience value of five epochs, was implemented to limit overfitting and avoid unnecessary training iterations. Prior to model training, all input variables were normalized using z-score standardization derived from the statistics of the training dataset.

3.4.2 Evaluation Metrics

Model performance was evaluated using a combination of continuous and categorical metrics. They are well suited for daily precipitation assessment, with particular emphasis on extreme and intermittent rainfall events. continuous evaluation metrics included the root mean square error (RMSE), which quantifies the overall magnitude of estimation errors; the correlation coefficient (CC), which measures the linear association between observed $(y_i)$and estimated $(\hat{y}_i)$ precipitation; and the mean error (ME), used to assess systematic bias. In addition, categorical metrics were applied to evaluate precipitation occurrence using a binary threshold of 0.1 mm. These included the POD, which measures the ability of the model to correctly identify observed precipitation events, and the false alarm ratio (FAR), which quantifies the proportion of incorrectly predicted events. Together, these complementary metrics provide a comprehensive assessment of model accuracy, bias, and event-detection skill. This is particularly important for hydrological applications in semi-arid regions, where precipitation is sparse, highly variable, and often dominated by short-duration events.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4}$$

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}} \tag{5}$$

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{6}$$

$$\mathrm{POD} = \frac{\mathrm{Hits}}{\mathrm{Hits} + \mathrm{Misses}} \tag{7}$$

$$\mathrm{FAR} = \frac{\mathrm{False\ Alarms}}{\mathrm{Hits} + \mathrm{False\ Alarms}} \tag{8}$$

Figure 3 presents the overall workflow of the proposed methodology, illustrating the main steps from data preprocessing to model development and evaluation for the DNN, LSTM, and CNN–LSTM architectures.

4. Results and Discussion

4.1 Comparison of Deep Learning Models

To evaluate the effectiveness of different deep learning architectures in estimating daily rainfall, the performance of DNN, LSTM, and CNN–LSTM models was systematically compared. The comparison was conducted across all synoptic stations and temporal aggregations. The results of different metrics for the models averaged across all the stations are illustrated as Figure 4.

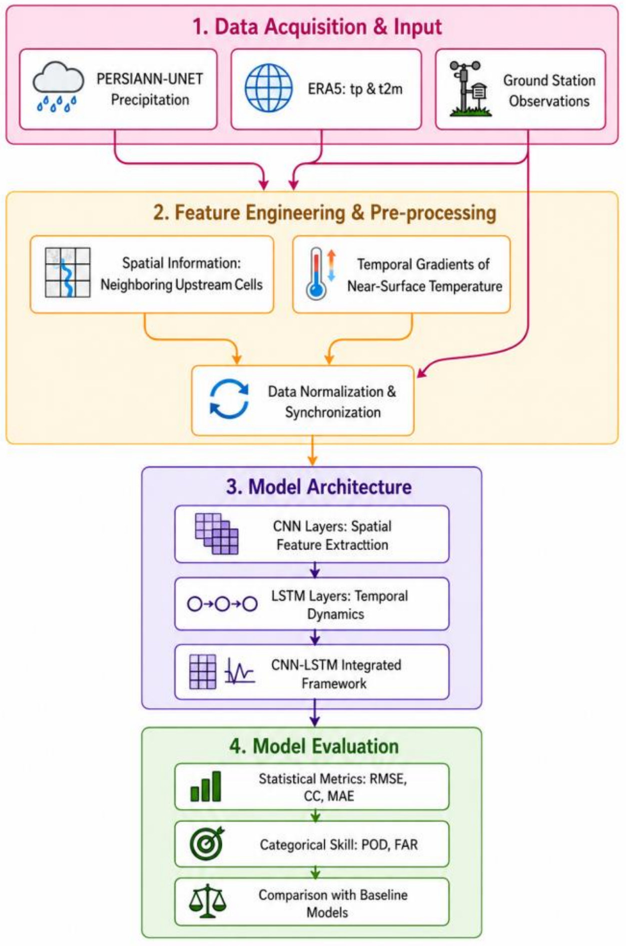

Fig. 3: Workflow of data processing, modeling, and evaluation

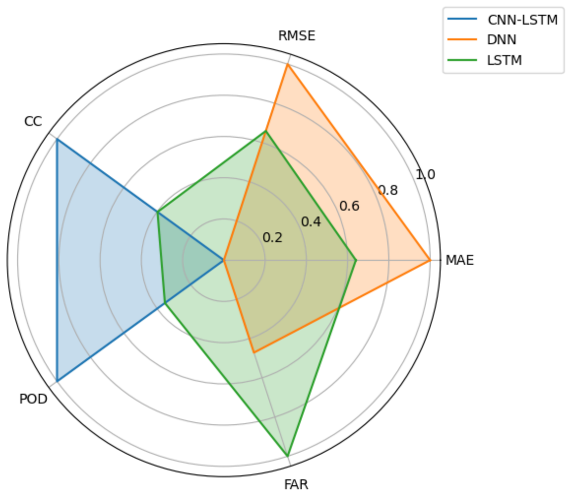

Fig. 4: Radar plot of the metrics per model

Figure 4 presents a radar chart comparing the performance of the three models (CNN-LSTM, DNN, and LSTM) using five evaluation metrics: MAE, RMSE, CC, POD, and FAR. Overall, the CNN–LSTM model exhibited the most balanced performance among the tested architectures. In particular, its higher CC and POD values indicates a stronger ability to represent rainfall variability and to correctly detect precipitation events across different conditions. In contrast, the DNN model exhibits the highest errors, with large MAE and RMSE values. It performs poorly in terms of POD, suggesting limited skill in accurately detecting rainfall occurrences. However, it shows moderate performance in FAR. The LSTM model shows intermediate performance. It produced lower errors than DNN and achieved the lowest FAR, indicating fewer false alarms; however, this improvement came with somewhat reduced CC and POD compared with the CNN–LSTM model. These results suggest that incorporating both spatial and temporal feature extraction provides a clear advantage for precipitation estimation. By combining convolutional and recurrent components, the CNN–LSTM architecture is able to reduce estimation errors more precisely while improving rainfall event detection. Consequently, it demonstrates greater suitability for rainfall estimation compared with the standalone DNN and LSTM models. To provide more comparison, Figure 5 illustrates boxplot of MAE metric for the models averaged over the stations.

Figure 5 presents the distribution of MAE values for the DNN, LSTM, and CNN-LSTM models using boxplots, allowing comparison of both the central tendency and variability of model errors. The DNN model exhibits the highest median MAE, approximately 1.2–1.3, along with a relatively wide interquartile range (IQR), indicating consistently larger errors and greater variability across samples. The LSTM model shows an improvement over DNN, with a lower median MAE of about 1.1 and a slightly reduced spread, although it still displays occasional high-error cases with upper values approaching 1.9. The CNN-LSTM model achieves the best performance, with the lowest median MAE near 1.0, a narrower IQR, and a reduced upper range (approximately 1.5), suggesting both improved accuracy and greater stability. Additionally, the lower whiskers for all models fall near 0.55–0.60, indicating similar best-case performance. However, the overall shift toward lower MAE values for the CNN–LSTM model confirms its superior error reduction capability. These results further support the effectiveness of the hybrid CNN–LSTM architecture. It produces more accurate and consistent rainfall predictions than the standalone DNN and LSTM models.

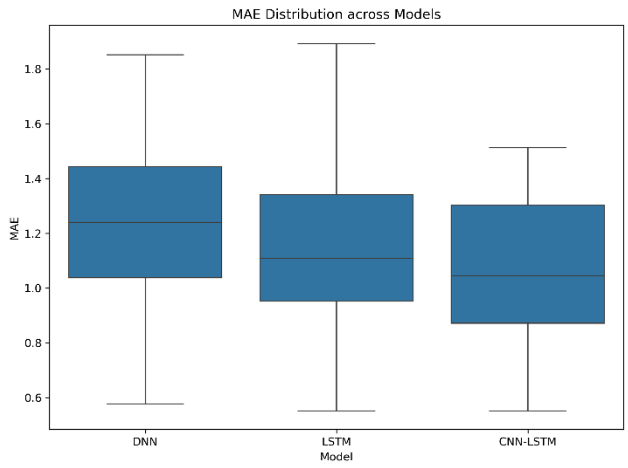

Fig. 5: Boxplot of the average MAE per model

4.2 Spatial Variability of Model Performance

Assessing how model performance changes across different stations is important for evaluating the robustness and transferability of rainfall prediction models under varying geographical and climatic conditions. Topography, local climate conditions, and station-specific rainfall patterns can strongly influence model accuracy. As a result, model performance may vary across different stations. Therefore, Evaluating errors at individual stations helps to reveal how effectively each model captures spatial differences in rainfall behavior. To examine this aspect, RMSE values were evaluated for each of the eight stations using the DNN, LSTM, and CNN–LSTM models. The results are presented in Figure 6 as a bar plot.

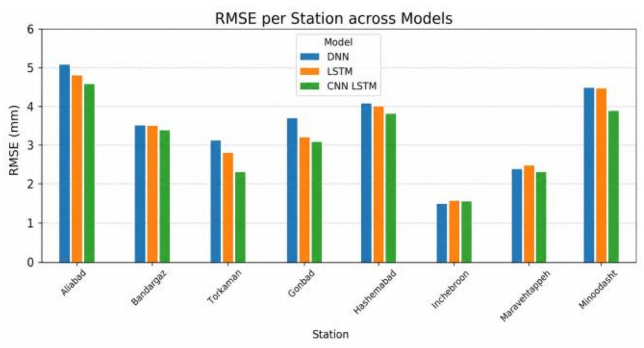

Fig. 6: Comparison of RMSE of rainfall estimates across the eight stations for the DNN, LSTM, and CNN-LSTM models

Figure 6 shows the RMSE values across the eight synoptic stations for the DNN, LSTM, and CNN–LSTM models. The CNN–LSTM model produces the lowest RMSE at seven of the eight stations, with values ranging from about 1.8 to 4.7 mm. This indicates a consistent reduction of prediction errors across stations with different geographic characteristics. The largest improvement is observed at Torkaman station, where the CNN–LSTM model reduces the RMSE to 2.3 mm, compared with 3.0 mm for the DNN model, corresponding to a 23% decrease. A similar trend appears at Minoodasht station, where the CNN–LSTM achieves an RMSE of 3.9 mm, while both the DNN and LSTM models yield 4.7 mm, representing an improvement of about 17%. These results suggest that explicitly incorporating spatial information from neighbouring cells helps the model better capture the systematic west–east precipitation propagation patterns. Notably, Inchebroon station presents a contrasting result. At this location, the DNN achieves the lowest RMSE (1.6 mm), outperforming both LSTM (1.8 mm) and CNN–LSTM (1.8 mm) models. This finding reinforces the earlier observation that simpler functional mappings may be more suitable here. The result likely reflects decoupled local rainfall generation mechanisms or dominant orographic forcing that overrides regional-scale advection.

In contrast, the highest absolute errors are observed at Aliabad station, where RMSE values range from approximately 4.7 to 5.2 mm across all models. This suggests persistent modeling challenges in this region. These difficulties may be associated with complex terrain interactions or greater precipitation variability that is not adequately resolved by the satellite and reanalysis inputs.

At Hashemabad and Maraveh-Tappeh, the differences among the models are minimal (less than 0.3 mm). This indicates that, at these stations, increasing model complexity beyond basic nonlinearity offers only marginal improvement. Overall, the RMSE analysis supports the correlation-based findings presented in Figure 6. It confirms that the spatiotemporal CNN–LSTM architecture generally reduces the magnitude of errors. At the same time, it remains flexible enough to capture station-specific exceptions. In these cases, a simpler model structure may be more appropriate.

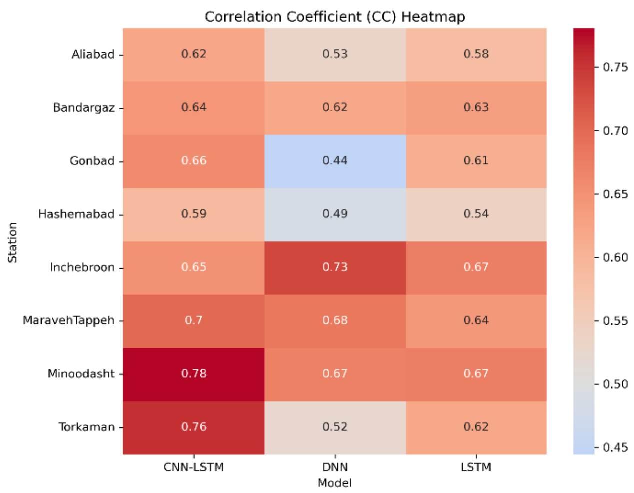

Fig. 7: Heatmap of the CC matrix for the DNN, LSTM, and CNN-LSTM models across the stations

Figure 7 presents the CC matrix, comparing the performance of DNN, LSTM, and CNN-LSTM models across eight synoptic stations in the Golestan Province. The CNN-LSTM model demonstrates superior performance at six of the eight stations, with CC values ranging from 0.59 to 0.78, achieving its highest correlation at Minoodasht station (CC = 0.78) and Torkaman station (CC = 0.76).

This improvement can be explained by the complementary structure of the model. The CNN layers extract spatial features from neighboring grid cells, while the LSTM layers capture temporal dependencies in the rainfall sequence. By combining these components, the CNN–LSTM model can better represent the west–east propagation patterns that characterize precipitation systems in the region. Notably, at Incheboron station, the DNN model unexpectedly achieves the highest CC of 0.73. This surpasses both the CNN–LSTM (0.65) and LSTM (0.67) models. This result suggests that, at this location, a simpler nonlinear relationships may dominate over complex spatiotemporal patterns. The finding is likely influenced by localized orographic effects or distinct microclimatic conditions. The LSTM model generally performs at an intermediate level compared with the DNN and CNN–LSTM models across most stations. This confirms that temporal memory alone, without explicit spatial context, provides only limited improvement. The DNN model shows the most variable performance, with its poorest result at Gonbad station (CC = 0.44). At this site, the absence of both spatial and temporal context severely limits predictive skill. Overall, these results highlight the importance of incorporating spatiotemporal dependencies for sub-daily precipitation bias correction in data-scarce regions. They also emphasize that model architecture should be adapted to site-specific characteristics when appropriate.

4.3 Influence of Input Features and Data Sources

The primary CNN-LSTM model was developed using all available input features. These include ERA5 total precipitation (tp), PERSIANN-UNet rainfall, near-surface air temperature (t2m), and the temporal gradient of t2m. This full input set was designed to capture both precipitation signals and key thermodynamic conditions related to rainfall formation. The scatter plots of predicted versus observed rainfall across all stations for this model are presented in Figure 8.

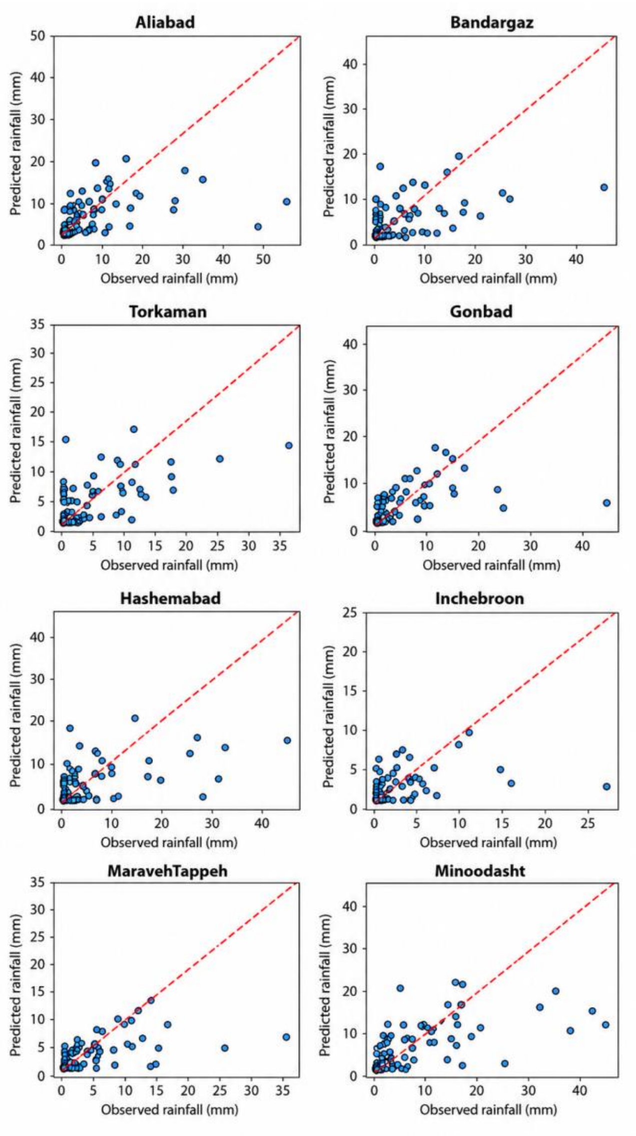

Fig. 8: Observed versus predicted rainfall scatter plots for the CNN-LSTM model with all input features

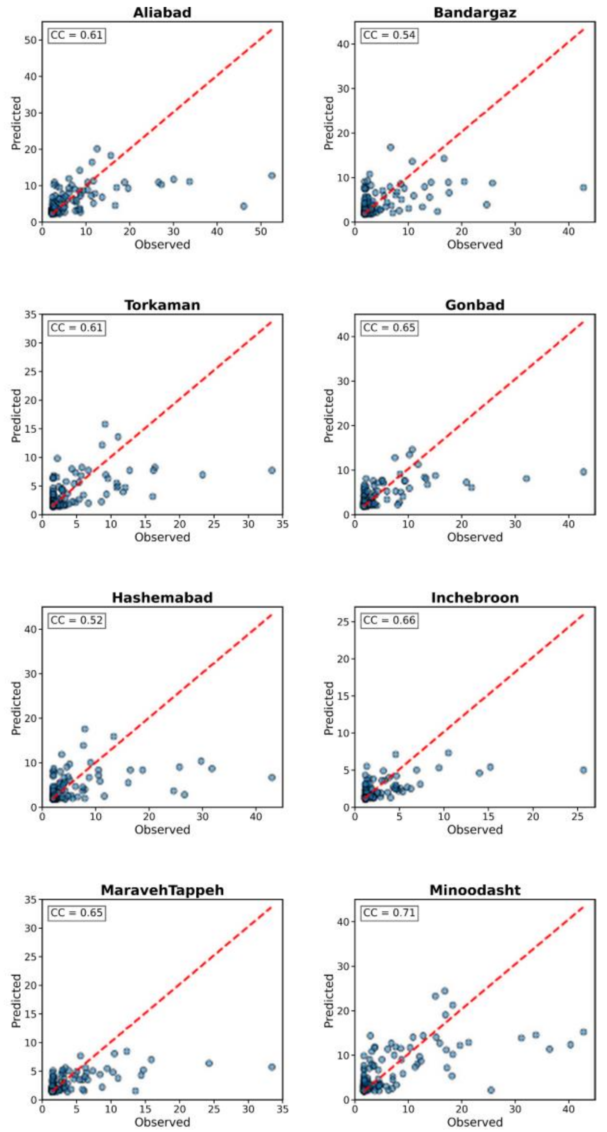

Fig. 9: Observed versus predicted rainfall scatter plots for the CNN-LSTM model excluding near-surface air temperature and its temporal gradient

To evaluate the importance of temperature-related features, an additional experiment was conducted. In this setup, t2m and its temporal gradient were removed from the input structure, while all other variables were kept unchanged. This approach allows the role of temperature information to be examined directly without modifying the overall model architecture. The scatter plots of observed versus predicted rainfall for the CNN-LSTM model excluding temperature-related inputs at all eight stations are shown in Figure 9.

A comparison between Figures 8 and 9 highlights the influence of temperature-related features on model performance across all eight stations. When all input features are included (Figure 8), the predicted rainfall values generally agree better with observations. The scatter points are closer to the 1:1 line, indicating higher accuracy. In contrast, the model excluding near-surface temperature and its temporal gradient (Figure 9) shows more scatter and larger deviations, particularly for extreme rainfall events.

For example, at the Bandartorkman station, an observed extreme rainfall of about 35 mm was predicted as roughly 15 mm by the full-input model, and only around 5 mm by the model excluding temperature. This illustrates that the absence of temperature-related features further reduces the model's ability to capture high-intensity events. Overall, the results across all stations demonstrate that including t2m and its temporal gradient improves model efficiency and provides valuable information for representing both typical and extreme rainfall events.

Generally, deep learning models tend to underestimate extreme rainfall events. Even with a weighted loss to address the imbalance in the data, the limited number of extreme samples causes the models to be biased toward the more frequent low or zero values. Consequently, accurately predicting rare, high-intensity events remains challenging, especially in regions with complex spatiotemporal rainfall patterns.

5. Conclusions

This study investigated the capability of deep learning models for daily rainfall estimation across eight synoptic stations in Golestan Province. Three architectures—DNN, LSTM, and CNN-LSTM—were compared using multiple evaluation metrics, including MAE, RMSE, CC, POD, and FAR. The role of input features was also examined, with a particular focus on near-surface air temperature (t2m) and its temporal gradient. The results show that combining convolutional and recurrent structures with carefully selected inputs, substantially improves rainfall prediction, both in overall accuracy and in representing extreme events.

- The CNN-LSTM model consistently outperformed DNN and LSTM models across most stations and metrics, showing the lowest errors and highest CCs.
- Incorporating both spatial (via CNN) and temporal (via LSTM) information enhances model skill, enabling more accurate rainfall estimation than using either structure alone.
- Temperature-related features (t2m and its temporal gradient) substantially improve the prediction of extreme rainfall events, as reflected by the reduced underestimation in high-intensity cases.
- The hybrid model shows robustness across diverse geographic conditions, with consistent performance even in stations with complex terrain or high rainfall variability.
- Removing temperature-related predictors results in a clear decline in model performance, highlighting the importance of thermodynamic information alongside precipitation data.
- Simpler models such as DNN or standalone LSTM may perform reasonably well at some stations. However, they generally struggle to capture spatiotemporal rainfall patterns, especially during extreme events or rapid rainfall transitions.

Overall, the proposed CNN-LSTM model demonstrates a high efficiency in estimating daily rainfall and capturing both typical and extreme events. However, the current study focused on estimation rather than forecasting, and the temporal resolution is limited to daily rainfalls. Future research should explore sub-daily rainfall prediction, which is crucial for flash flood early warning and hydrological modeling. Future studies could test the CNN–LSTM model across multiple basins with different rainfall behaviors. This would help assess its generalizability. The model could also include additional data sources, such as IMERG, GPM, or local radar. Using higher-resolution datasets would allow evaluation of its performance in forecasting and extreme events.

References

[1] Center for Hydrometeorology and Remote Sensing (CHRS). CHRS Data Portal: PERSIANN-UNet precipitation product. University of California, Irvine. Available at: https://chrsdata.eng.uci.edu/.

[2] Copernicus Climate Change Service (C3S). ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Data Store; 2018.

[3] Wang, F., Tian, D., & Carroll, M. (2023). Customized deep learning for precipitation bias correction and downscaling. Geoscientific Model Development, 16(2), 535–556.

[4] Apaza-Vilca, Cristhian, Maria Liz Mamani-Yupanqui, and Efrain Lujano. "Performance Evaluation of the ERA5, MERRA-2, and PERSIANN-CDR Gridded Products in the Tambo Basin." Environmental and Earth Sciences Proceedings 32.1 (2025): 17.

[5] Putro, Danurendro Ramadhani, and Obaja Triputera Wijaya. "Evaluating PERSIANN-CDR and ERA5 Precipitation Products: Insights from Bandung Raya." Jurnal Teknik Sipil 25.1 (2025): 1623-1633.

[6] Saravanan, Aatralarasi, et al. "Evaluation of remote-sensing-and reanalysis-based precipitation products for agro-hydrological studies in the semi-arid tropics of Tamil Nadu." Hydrology and Earth System Sciences 29.19 (2025): 4847-4870.

[7] Bagiliko, John, et al. "Validation of satellite and reanalysis rainfall products against rain gauge observations in Ghana and Zambia." Theoretical and Applied Climatology 156.5 (2025): 1-20.

[8] Assamnew, Abera Debebe, Birhan Gessese Gobie, and Birhanu Asmerom Habtemicheal. "Evaluating the accuracy of gridded precipitation datasets and CMIP6 GCM ensemble means against gauge observations in South Wollo, Ethiopia." Scientific Reports 15.1 (2025): 34197.

[9] Zhang, Jianbin, Zhiqiu Gao, and Yubin Li. "Deep-learning correction methods for weather research and forecasting (WRF) model precipitation forecasting: a case study over zhengzhou, China." Atmosphere 15.6 (2024): 631.

[10] Hu, Yi-Fan, Fu-Kang Yin, and Wei-Min Zhang. "Deep learning-based precipitation bias correction approach for Yin–He global spectral model." Meteorological Applications 28.5 (2021): e2032.

[11] Miri, S. M., M. R. Kavianpour, and M. J. Alizadeh. "Short-term rainfall forecasting using multi-task learning and Weibull based postprocessing technique." International Journal of Environmental Science and Technology 22.14 (2025): 13571-13584.

[12] Yang, Xiaoying, et al. "Correcting the bias of daily satellite precipitation estimates in tropical regions using deep neural network." Journal of Hydrology 608 (2022): 127656.

[13] Moraux, Arthur, et al. "A deep learning multimodal method for precipitation estimation." Remote Sensing 13.16 (2021): 3278.

[14] Tao, Yumeng, et al. "A deep neural network modeling framework to reduce bias in satellite precipitation products." *Journal of Hydrometeorology* 17.3 (2016): 931-945.

[15] Akbari Asanjan, Ata, et al. "Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks." *Journal of Geophysical Research: Atmospheres* 123.22 (2018): 12-543.

[16] Sadeghi, Mojtaba, et al. "Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information." *Environmental Modelling & Software* 134 (2020): 104856.

[17] Amini, Amirmasoud, Mehri Dolatshahi, and Reza Kerachian. "Effects of automatic hyperparameter tuning on the performance of multi-variate deep learning-based rainfall nowcasting." *Water Resources Research* 59.1 (2023): e2022WR032789.

[18] Amini, Amirmasoud, Mehri Dolatshahi, and Reza Kerachian. "Real-time rainfall and runoff prediction by integrating BC-MODWT and automatically-tuned DNNs: Comparing different deep learning models." *Journal of Hydrology* 631 (2024): 130804.

[19] Aderyani, Fatemeh Rezaei, S. Jamshid Mousavi, and Fatemeh Jafari. "Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN." *Journal of Hydrology* 614 (2022): 128463.

[20] Li, Peifeng, Jin Zhang, and Peter Krebs. "Prediction of flow based on a CNN-LSTM combined deep learning approach." Water 14, no. 6 (2022): 993.

[21] Ahmed, AA Masrur, Shahida Akther, Thong Nguyen-Huy, Nawin Raj, S. Janifer Jabin Jui, and S. Z. Farzana. "Real-time prediction of the week-ahead flood index using hybrid deep learning algorithms with synoptic climate mode indices." Journal of Hydro-Environment Research 57 (2024): 12-26.

[22] Wu, H., Yang, Q., Liu, J., & Wang, G. (2020). A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. *Journal of Hydrology*, *584*, 124664.

[23] Ishida, K., Ercan, A., Nagasato, T., Kiyama, M., & Amagasaki, M. (2024). Use of one-dimensional CNN for input data size reduction in LSTM for improved computational efficiency and accuracy in hourly rainfall-runoff modeling. *Journal of Environmental Management*, *359*, 120931.

[24] Gavahi, K., Foroumandi, E., & Moradkhani, H. (2023). A deep learning-based framework for multi-source precipitation fusion. *Remote Sensing of Environment*, *295*, 113723.

[25] Mousavimehr, SM., & Kavianpour, MR. (2025). A non-stationary downscaling and gap-filling approach for GRACE/GRACE-FO data under climatic and anthropogenic influences. *Applied Water Science*, 15, 91.

[26] Alizadeh, MJ., & Nourani, V. (2024). Multivariate GRU and LSTM models for wave forecasting and hindcasting in the southern Caspian Sea. *Ocean Engineering*, *298, 117193.*

[27] Ko, J., Lee, K., Hwang, H., Oh, S. G., Son, S. W., & Shin, K. (2022). Effective training strategies for deep-learning-based precipitation nowcasting and estimation. *Computers & Geosciences*, *161*, 105072.