Jurnal Sumber Daya Air
Vol. 22 No. 1, Mei 2026

P- ISSN 1907 - 0276
E- ISSN 2548 - 494
Terakreditasi Kemenristek Dikti
http://jurnalsda.pusair-pu.go.id

Pemodelan Hidrograf Berbasis LSTM (*Long Short-Term Memory*) untuk Waduk Pandanduri, Indonesia

LSTM (Long Short-Term Memory)-based Hydrograph Modeling for Pandanduri Reservoir, Indonesia

Piter Wongso1) Siska Wulandari2) Neil Andika3)* Faizal Immaddudin Wira Rohmat1) Muhammad Ammar Fadhil Adfa4)

1) Bandung Institute of Technology, Jalan Ganesa No. 10, Bandung 40132, West Java, Indonesia
2)Department of Integrated Water Management Engineering, Institut Teknologi Sumatera, South Lampung 35365, Lampung, Indonesia
3)Department of Civil and Environmental Engineering, Gadjah Mada University, Yogyakarta, Indonesia
4)Ministry of Public Works, Republic of Indonesia, JL. Pattimura 20, Kebayoran Baru Jakarta,12110 Indonesia
**Corresponding author*: faizalrohmat@itb.ac.id

Received: 6 October 2025; Revised: 14 March 2026; Accepted: 29 May 2026

Abstract

Dams are vital infrastructure that supports irrigation, raw water supply, flood control, hydroelectric power, and tourism. Reservoirs of the dam store the water during the rainy season and release it in the dry season to reduce the flood risk. Continuous monitoring is essential to ensure their optimal function and safe operation. However, in many developing regions, the availability and quality of discharge data remain inadequate due to poor data governance, limited resources, and insufficient digital infrastructure. These challenges highlight the need for new approaches capable of improving discharge prediction under limited data conditions. To address this issue, Machine Learning (ML) has gained popularity in hydrological research. Long Short-Term Memory (LSTM) is one of the ML methods that has proven effective for analyzing historical data patterns for prediction. This study uses an LSTM-based hydrograph model in the Pandanduri Reservoir Watershed, Indonesia. Compared to conventional statistical or conceptual models, LSTM offers the advantage of capturing nonlinear dynamics in long time-series datasets. The model utilized open data, including satellite rainfall data, meteorological data, and streamflow data records, to capture the complex relationship between rainfall and streamflow. The generated model was validated with the recorded streamflow data. The model evaluation produces an NSE of 0.44 and an RMSE of 0.74. These values indicate that the model is capable of reproducing key rainfall–runoff dynamics under data-limited conditions. Further research is essential to improve the ability of LSTM to capture extreme events and improve its generalization across hydrological conditions.

Keywords: *hydrograph model, LSTM, machine learning, pandanduri reservoir, satellite data*

Abstrak

Bendungan merupakan infrastruktur vital yang berfungsi untuk mendukung irigasi, penyediaan air baku, pengendalian banjir, pembangkit listrik tenaga air, dan pariwisata. Waduk menyimpan air pada musim hujan dan melepaskannya pada musim kemarau untuk mengurangi risiko banjir. Pemantauan secara kontinu sangat penting untuk memastikan fungsi tetap optimal dan operasional bendungan yang aman. Namun, di banyak wilayah berkembang, ketersediaan dan kualitas data debit masih tidak cukup akibat tata kelola data yang kurang baik, keterbatasan sumber daya, dan infrastruktur digital yang belum memadai. Tantangan ini menunjukan perlunya pendekatan baru yang mampu untuk meningkatkan prediksi debit pada kondisi keterbatasan data. Dalam bidang hidrologi, penerapan Machine Learning (ML) semakin populer, khususnya metode Long Short-Term Memory (LSTM) yang terbukti efektif dalam menganalisis pola data historis untuk melakukan prediksi. Penelitian ini menggunakan model hidrograf berbasis LSTM pada DAS Waduk Pandanduri, Indonesia. Dibandingkan dengan model statistik atau model konvensional, LSTM memiliki keunggulan dalam memperoleh dinamika nonlinier pada data dengan waktu yang panjang. Model ini memanfaatkan data yang tersedia secara bebas, termasuk data curah hujan satelit, data meteorologi, dan data debit sungai observasi untuk menangkap hubungan kompleks antara hujan dan aliran sungai. Model kemudian divalidasi dengan data debit observasi. Evaluasi model menghasilkan nilai NSE sebesar 0,44 dan RMSE sebesar 0,74. Nilai ini menunjukkan bahwa model mampu merepresentasikan hubungan hujan–aliran meskipun pada kondisi keterbatasan data. Penelitian lanjutan diperlukan untuk meningkatkan kemampuan LSTM dalam menangkap kejadian ekstrem serta meningkatkan generalisasi model pada berbagai kondisi hidrologi.

Kata Kunci*: model hidrograf, LSTM, pembelajaran mesin, waduk pandanduri, data satelit*

DOI: https://doi.org/10.32679/jsda.v22i1.1059

INTRODUCTION

A dam is a multifunctional structure that provides numerous benefits. Together with its associated water storage system, the reservoir, a dam supports irrigation, supplies raw water, controls flood, generates hydroelectric power, and offers opportunities for tourism. Effective monitoring of both the dam and the reservoir is crucial, particularly the monitoring of water inflow data. Inflow data plays a vital role in reservoir management, especially in supporting operational decision-making (Sushanth et al., 2023). However, in many developing regions, the availability and quality of inflow data remain inadequate (Biswas et al., 2022; Taye et al., 2023). This challenge arises primarily from issues such as poor data governance, limited resources, and insufficient digital infrastructure. To address these limitations, it is critical to adopt and implement solutions that enhance inflow prediction and improve reservoir management practices.

For decades, conventional methods include empirical models, conceptual models, and physically based models, have been effectively used to support hydrological modeling; however, hydrological uncertainties remain a challenge, especially in dealing with non-linear flows in a complex environments (Maurya, 2024). Considering the importance of dam inflow data, adaptations of model that can address these challenges are necessary– especially in the context of limited observational data and high variability of the rainfall-runoff relationship.

In response to these limitations, the use of Machine Learning (ML) and Artificial Intelligence (AI) has gained popularity in hydrology. ML has been applied in various aspects of hydrology, such as flood forecasting and prediction, river water quality modeling, and drought assessment, offering a more comprehensive and adaptive approach than conventional methods (Burnama et al., 2023; Zounemat-Kermani et al., 2021). ML also provides several advantages, such as higher predictive accuracy (Pathan et al., 2024), the ability to handle or to work in complex data sets (Hasan et al., 2024), and the ability to improve the interpretability of the results. Furthermore, ML allows for faster and quicker response to changes in hydrological conditions. Besides, integrating ML with automation systems can enhance the decision-making processes and improve reservoir operations. Therefore, with these various advantages, the adopting of ML for streamflow prediction in reservoirs has a great possibility for more effective reservoir management.

Long Short-Term Memory (LSTM) is an ML technique that has shown effectiveness in the field of hydrology. The LSTM in ML is developed from the RNN structure. This method adds memory cell units to the hidden layer, replacing the original cells in the RNN, thus overcoming several weaknesses of standard RNNs, one of which is the vanishing gradient, which is the loss of gradients in the long sequence training process (Sahu et al., 2023). LSTM with a gate mechanism helps maintain gradients longer. Short-term memory allows previous input information to be stored in the network, which affects the output (Yadav et al., 2023). Some of the advantages of LSTM that support the modeling of complex relationships in hydrology include: the capability to investigate relationships between historical data and predictions (Andika et al., 2025; Enung et al., 2022; Kardhana et al., 2022); proficiency in predicting multiple hydrological variables (Z. Xiang et al., 2020); excellence in hydrological simulation across large datasets (Lees et al., 2021); outperforms the performance compared to the conventional model when calibrated for discharge (streamflow) prediction across diverse climate conditions (de Moura et al., 2022); and its enhanced accuracy in simulating complex hydrological systems (Li et al., 2022).

This study aims to develop an LSTM-based hydrograph model for the Pandanduri Reservoir River Basin in Indonesia. The Pandanduri Reservoir, illustrated in Figure 1 is situated in the Palung watershed in the Sakra district of Lombok Timur Regency, in the Nusa Tenggara Barat Province of Indonesia (Aribowo et al., 2023). This Reservoir encompasses a watershed area of 64.51 km^2 (Trimartinni et al., 2024) and has a total volume of 27.20 million m^3 with an inundated area spanning 315 ha. It serves multipurpose, including irrigation (5,168 ha) and raw water supply (50 L/s), and benefits such as flood control, inland fisheries, tourism and the development of micro-hydroelectric generators (BBWS, 2025). The average annual rainfall in the area is 1,899 mm, with an annual influx of 75.57 million m^3 from the Suradadi River, feeding into the Pandanduri Reservoir.

Study Site

The Pandanduri Reservoir was selected as the study site due to the limited application of machine learning modeling in hydrology for the reservoirs in Lombok. Besides that, the Pandanduri reservoir has observational streamflow data available over a time span that allows for hydrological modeling using the LSTM. The data used throughout the modeling timeframe must cover the wet and dry seasons, annual climate variations, and the probability of extreme events to achieve more accurate modeling results. This is because it can reliably capture the dynamics of the hydrological system at the study site (Majone et al., 2022; Snider et al., 2024). Moreover, the model utilizes open data sources, including satellite rainfall data from GPM (Global Precipitation Measurement), meteorological data from the Meteorology, Climatology, and Geophysics Agency of Indonesia (BMKG), and streamflow data from the Indonesia Ministry of Public Works (PU), to accurately capture the complex nonlinear relationship between rainfall and streamflow in the Pandanduri Reservoir. The developed model is expected to help engineers perform hydrological analysis in the Reservoir more quickly, efficiently, and accurately.

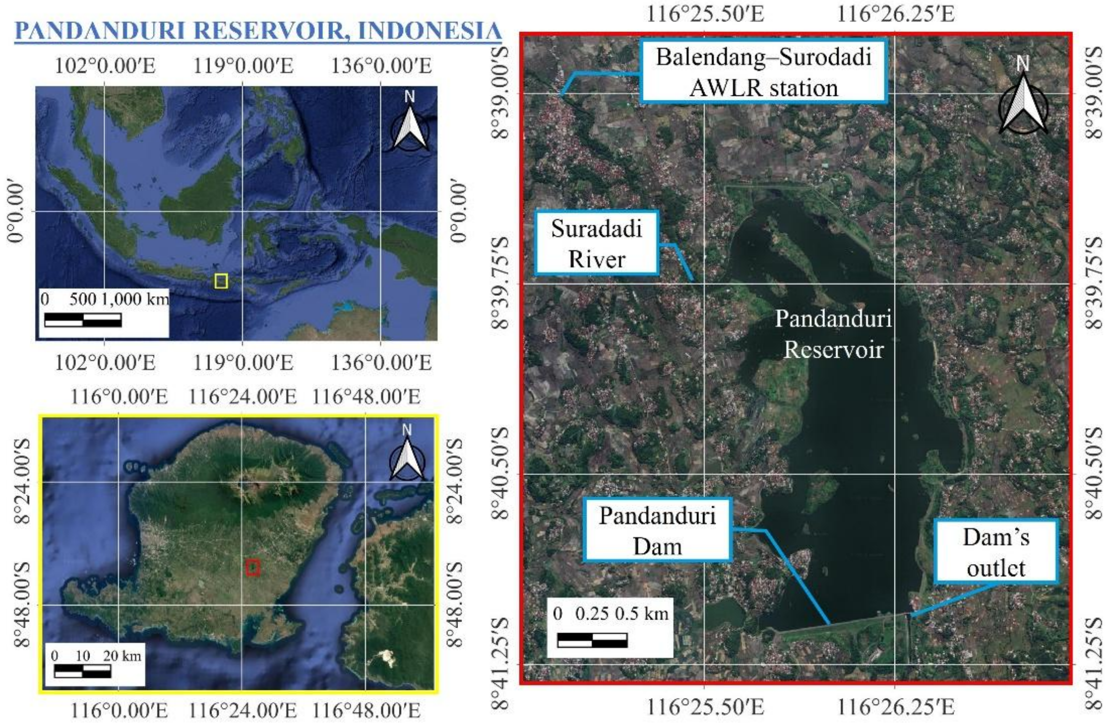


Figure 1 Pandanduri Reservoir is in West Nusa Tenggara Province, Indonesia, specifically on the island of Lombok (highlighted in yellow). The figure highlighted in red shows the current situation of the Pandanduri Reservoir.

METHODS

Flowchart

The flowchart (Figure 2) of this research outlines several key steps. The process commences with the gathering of three main types of input data: daily rainfall, evapotranspiration (obtained from processed meteorological data), and streamflow data. The dataset then underwent an evaluation and correction process. Following this, the flowchart diverges into two separate paths: the training dataset and the testing dataset. The training dataset is utilized to develop the initial model, an untrained model which is then transformed into a trained LSTM model. The testing set is used to evaluate the predictive value or the accuracy of the trained LSTM model. In the end, the trained LSTM model produces daily streamflow predictions, which are then compared againt the actual or recorded streamflow data. The best model selected is the model that demonstrated the best Root Mean Square Error (RMSE) and Nash-Sutcliffe Efficiency (NSE) values.

The recorded streamflow hydrograph data of the Pandanduri watershed contain three notable inaccuracies (errors) in the historical record. To address the inherent bias found in this raw data, a technique for bias correction was utilized to clean the data (Michael et al., 2022). To obtain the corrected daily average streamflow data for the study period, this study extrapolated the Julian date calendar approach. The Weighted Regressions on Time, Discharge, and Season (WRTDS) technique (Hirsch et al., 2010) is considered an alternative approach that offers a thorough framework for correcting biases. This technique constructs local regressions for daily estimates using three-dimensional weighting (time, discharge, and season). The integration with the Julian calendar allows for extrapolation to days for which no data (biases) are available, while accurately and periodically accounting for the seasonal position, ensuring that the flow data remains representative because it can capture seasonal patterns at the study site.

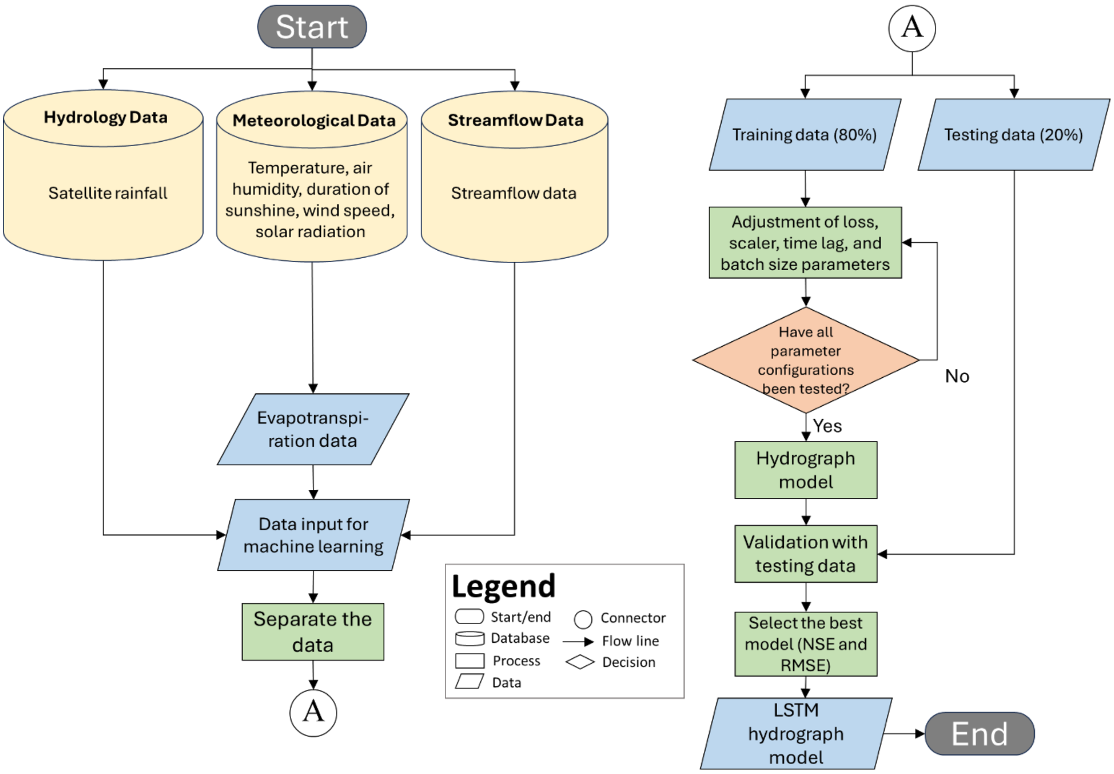


Figure 2 Flowchart research of hydrograph modeling using LSTM

Data Collection

The data needed to generate the hydrograph model consists of three types of data: rainfall data, meteorological data, and streamflow data. These three data are open data and proprietary data obtained from institutions in Indonesia, such as PU and BMKG. Data type, along with the description, specification, and sources, are briefly presented in Table 1.

Table 1 Data collection and preparation for hydrograph modeling

Data Type	Data Description	Specifications	Sources
Rainfall data	Satellite rainfall data	Daily, 2006–2013 (mm/day)	GPM (NASA, 2018)
Meteorological data	Climatic conditions in the study area generated evapotranspiration data	Daily, 2006-2013 (temperature (°C), relative humidity (%), wind speed (m/s), solar radiation (MJ/m^2/day))	BMKG, Zainuddin Abdul Madjid meteorological station
Streamflow data	Transformed from water level data	Daily, 2006–2013 (m^3/s)	PU, Balendang Surodadi *AWLR station

**AWLR=Automatic Water Level Recorder*

1. Rainfall Data

The rainfall data used in this study were sourced from satellite GPM data with a daily temporal resolution and a spatial resolution of 0.1° × 0.1° (NASA, 2018). This resolution provides sufficient variations of rainfall pixels, suitable for regional scale analysis. In this study, there are six GPM cells/grids covering the Pandanduri Reservoir watershed. To extract regional rainfall data, regional averaging was used to obtain a single daily regional rainfall value.

The GPM was developed by the Japanese Aerospace Exploration Agency (JAXA) and has been available since 2014. The data is integrated within the IMERG (Integrated Multi-satellite Retrievals for GPM) program owned by the National Aeronautics and Space Administration (NASA). As the successor product of the previous satellite rainfall measurement product, the TRMM (Tropical Rainfall Measuring Mission), GPM offers significant enhancement over TRMM's ability to detect precipitation in tropical regions, further augmented with the DPR (Dual-frequency Precipitation Radar) and GMI (GPM Microwave Imager) payloads. These advancements significantly lead to the better detection of solid and micro precipitation (0-1 mm/day). Additionally, GPM precipitation data offers greater precision in global satellite rainfall observations for hydrological research compared to the TRMM era (Draper et al., 2015). These advantages make the GPM data set is suitable choice for this study.

2. Meteorological Data

Meteorological data is used to classify climate conditions in an area with evapotranspiration (ET) values. Evapotranspiration, together with precipitation (rainfall), represents the climate of an area and is used to support decisions on agricultural water management and irrigation. A popular ET_0 calculation method is the Penman-Monteith method, a development of the Penman method recommended by FAO (K. Xiang et al., 2020). The variables considered are the duration of sunshine (hours), solar radiation, wind speed, air humidity, and temperature. All data were obtained from the Zainuddin Abdul Madjid meteorological station, which is the closest station to the Pandanduri Watershed (8°45'10.0" S 116°14'59.4" E).

$$ET_0 = \frac{0.408\Delta(R_n) + \gamma \frac{900}{T+273} U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where ET_0 is evapotranspiration (mm/day), R_n is the amount of radiation (MJ/m^2. day), U_2 is the wind speed (m/s), Δ is the slope of the saturation vapor pressure curve (kPa/°C); γ is the psychrometric constant; e_s is the saturation vapor pressure (kPa); and e_a is the actual vapor pressure (kPa).

3. Streamflow Data

Streamflow data were collected from the AWLR gauges, which is managed by PU, specifically carried out by the Hydrology Center, Research and Development Agency for Water Resources. The AWLR station, Balendang–Surodadi, is located at 8°39'03.1" S 116°24'56.0" E, situated near the inlet of the Pandanduri Reservoir (Figure 1) with the main River is the Palung River. The location of the AWLR in this upstream reservoir represents the accumulation of air from the entire overlying watershed entering the reservoir system as inflow, then allowing only the discharge entering the reservoir to be calculated without being affected by reservoir operational policies. The water level data were recorded at daily temporal resolution and were available from 1986 to 2013, although there were gaps in some years. The longest continuous dataset spans from 2006 to 2013, covering a total of 8 years, and this dataset is utilized in this study.

Machine Learning Processes

The LSTM method in this study combines rainfall and evapotranspiration data to create a hydrograph model. The LSTM model network is separated into two partial processes, with 80% of the dataset allocated for training while the remaining 20% is reserved for testing. The training data was taken in the range of 2006–2011, and the rest of the testing data was in the range of 2012–2013.

Adjustments are made to the loss, scaler, lag time, and batch size parameters to improve the model accuracy (Wang et al., 2022) (see Table 2 for modeling parameters and Figure 1 for the flowchart). The loss parameters in the model use the Huber and MSE methods. This parameter plays an important role in improving ML performance. The Huber loss function combines square loss and absolute loss to reduce the influence of outliers and prevent overfitting. The use of square loss within certain limits helps prevent too rapid gradient reduction and loss of optimum values (Wang et al., 2022). The loss equation is shown in equation (2), which δ in the equation represents the threshold parameter in the Huber loss function.

Table 2 Modeling Parameters

Component	Parameters
Input	Rainfall, ET_0
Optimization (Processes)	Loss, scaler, lag time, and batch size
Output	Streamflow (hydrograph)

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}\left(y - f(x)\right)^2, & if\ |y - f(x)| \leq \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & otherwise \end{cases} \quad (2)$$

$$, where\ parameter\ \delta > 0$$

ML utilized a series of data with various variable approaches to predict certain values. However, long data series generally have data bias. To prevent this bias, data normalization is needed. Data normalization is a technique used in data preprocessing that adjusts or scales the data (Singh & Singh, 2020), which is important for enhancing the quality of the model and accelerating its ability to estimate value accurately. This model used the scaler method for data normalization (Sharma, 2022), incorporating both the min-max and the robust methods. The min-max method establishes linearly determined lower and upper limits, usually scaling the data within the range of 0 to 1 or -1 to 1. The scaler equation is presented in equation (3).

$$x'_{i,n} = \frac{x_{i,n} - \min(x_i)}{\max(x_i) - \min(x_i)}(nMax - nMin) + nMin \quad (3)$$

The lag time is the required time between peak rainfall and peak streamflow. Lag time greatly affects peak streamflow, so it needs to be studied in flood forecasting (Black et al., 2021). In this study, lag time is analyzed across a range of values: 7, 30, 90, 180, and 365 days. The last parameter is the batch size, which is the number of data samples processed by the model in each iteration during training. The batch sizes used in this study were 4, 8, 16, 32, and 64. The effect of batch size will be the accuracy of the model and the time required to converge. Small batches can converge faster, while larger batches may achieve better performance but reduce updates for convergence. Furthermore, a higher learning rate tends to favor larger batch sizes, enhancing performance than a smaller rate (Kandel & Castelli, 2020).

RESULTS AND DISCUSSION

Hydrograph Modeling using the LSTM Method

Modeling was carried out using the AMD Ryzen 9 7950X 16-core Processor (32 CPU threads) at around 4.5GHz. The model was trained with 200 epochs, which have been determined to be the appropriate value so that it gave satisfactory results (Kratzert et al., 2018a; Nguyen et al., 2025). The model training process was fast, where the training time or 'training time' only took seconds to produce a hydrograph. Although using a large dataset, utilizing ML provides the advantage of being able to estimate hydrographs faster so that it can be used for any water resource analysis.

Parameter Analysis

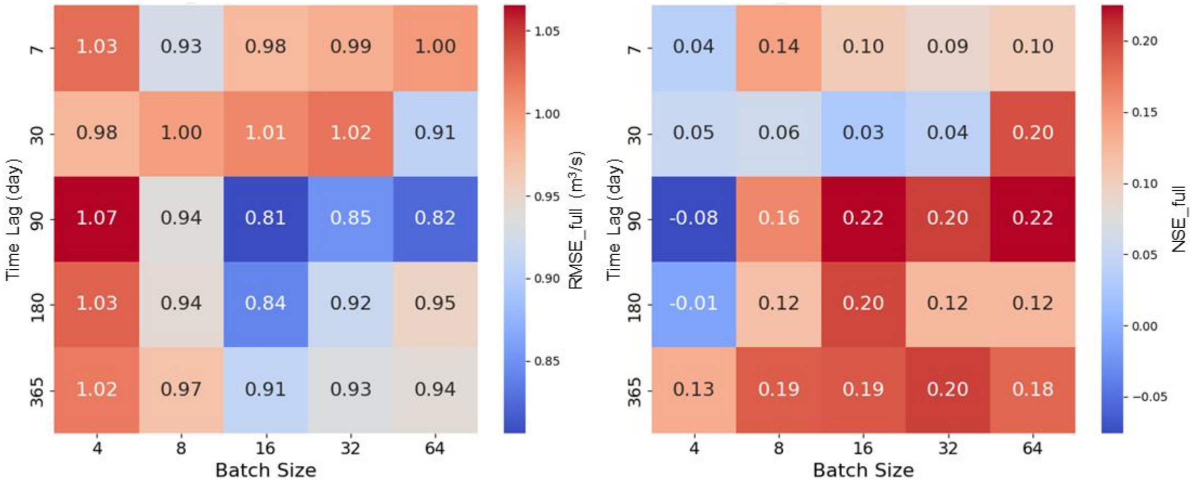
Differences in the parameters of lag time, loss method, scaler, and batch size resulted in 100 combinations in the results of this LSTM model. From all combinations of training experiments with changes in four different parameters, the model was evaluated using 100% of the dataset for 8 years, which was then compared to the streamflow data in 2006-2013. Figure 3 shows the matrix between two hyperparameters, batch size and lag time, as part of the parameters compared with the RMSE and NSE levels.

In the RMSE matrix (Figure 3a), a lower RMSE value indicates a better combination of batch size and lag time, which results in a model that is closer to the test model. Conversely, a higher NSE value (Figure 3b) indicates a better combination model. Various combinations of these parameters were evaluated, resulting in RMSE ranging from 0.81 to 1.07 and NSE values ranging from −0.08 to 0.22. Matrix analysis identified that the combination of batch size 16 and a lag time of 90 days gave the best performance, while the combination of batch size 4 and a lag time of 90 days gave the worst performance. This concludes that in this model, larger and more precise lag values do not always result in better model performance. Matrix analysis also shows differences in model performance. Models using small batch sizes with a lag time of 7 days consistently show higher errors, indicating that this combination is less effective in capturing hydrological dynamics. In contrast, the second group of models, with batch sizes of 16 to 64 and lag time between 90 and 365 days, shows superior performance with minimal errors. The results of this study emphasize the need to select appropriate hyperparameter values to optimize model accuracy and generalization capabilities. The observed pattern, in which larger batch sizes (16–64) consistently outperform smaller ones (4–8), aligns with findings from hydrological LSTM applications. Moderate batch sizes yield more stable gradient estimates for rainfall–runoff time series, particularly in the presence of seasonal patterns (Kratzert et al., 2018b). The superior performance of a batch size of 16 in this study likely reflects the annual hydrological cycle of the Pandanduri catchment, where 16 samples capture meaningful within-season variability without introducing excessive noise.

Figures 4a and b compare two scaler methods, robust and min-max. The results show that the robust scaler has a lower RMSE value, indicating better model fit and higher NSE compared to the min-max scaler. Similarly, Figures 4c and d examine the impact of MSE and Huber loss functions. The observed trend shows that the Huber loss function produces lower RMSE and higher NSE values compared to the MSE loss function. These findings highlight the important role of data preprocessing and loss function selection in optimizing model performance metrics. Although the final performance range remains consistent across methods, specific choices in lag time, batch size, scaler, and loss function can significantly affect RMSE and NSE values. This study identifies the optimal parameter combination with batch size 16, lag time 90 days, robust scaler, and Huber loss function, providing valuable insights for practitioners who want to develop accurate and reliable hydrological models.

Proposed Model, Training, and Model Validation

The learning curve (Figure 5a) shows a rapid decrease in both training and validation losses in the beginning, followed by a more gradual convergence toward zero. An increase in epochs results in smaller losses, indicating effective model learning. However, a divergence begins to occur between the validation and training losses starting from epoch 141, suggesting potential overfitting. To address overfitting, applying regularization techniques or early stopping strategies is essential (Srivastava et al., 2014). Nevertheless, the overall downward trend in loss makes this concern less critical, even though it leads to longer processing times. Figure 5b illustrates that the Mean Absolute Error (MAE) aligns with this trend, showing minimal difference between the training and validation metrics in the final epochs. By carefully selecting hyperparameters and incorporating appropriate regularization techniques, the models can achieve an improved generalization, leading to more reliable and accurate predictions.



Time Lag (day)	4	8	16	32	64
7	1.03	0.93	0.98	0.99	1.00
30	0.98	1.00	1.01	1.02	0.91
90	1.07	0.94	0.81	0.85	0.82
180	1.03	0.94	0.84	0.92	0.95
365	1.02	0.97	0.91	0.93	0.94

Time Lag (day)	4	8	16	32	64
7	0.04	0.14	0.10	0.09	0.10
30	0.05	0.06	0.03	0.04	0.20
90	-0.08	0.16	0.22	0.20	0.22
180	-0.01	0.12	0.20	0.12	0.12
365	0.13	0.19	0.19	0.20	0.18

Figure 3 Heatmap results of (a) RMSE and (b) NSE values regarding LSTM testing results from the testing process comparing all combination data sets.

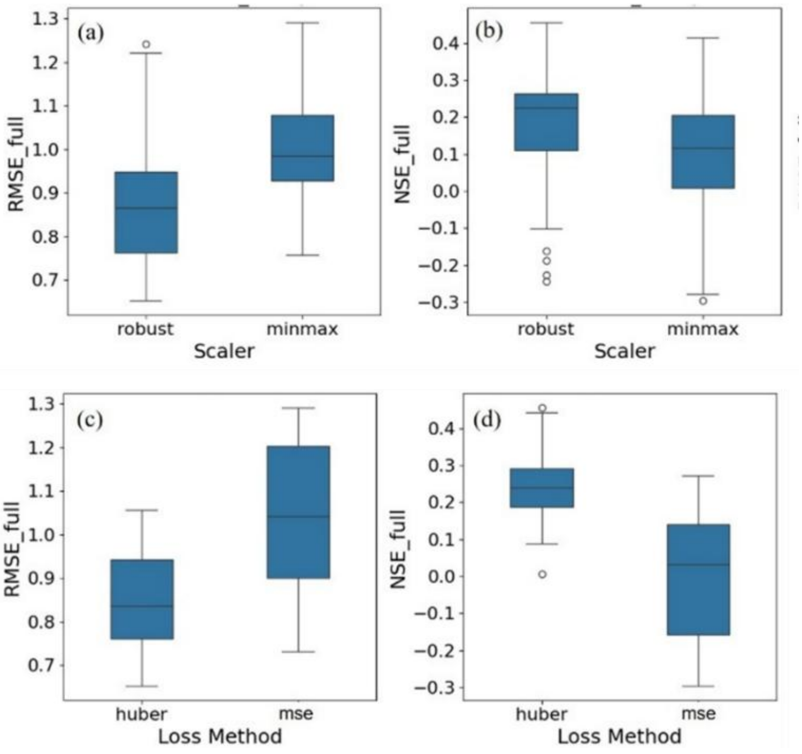


Figure 4 Box Plot Results of comparison of average RMSE (a) and NSE (b) comparison of min-max and robust variables on the scaler method; comparison of Average RMSE (c) and NSE (d) on the MSE and Huber loss methods.

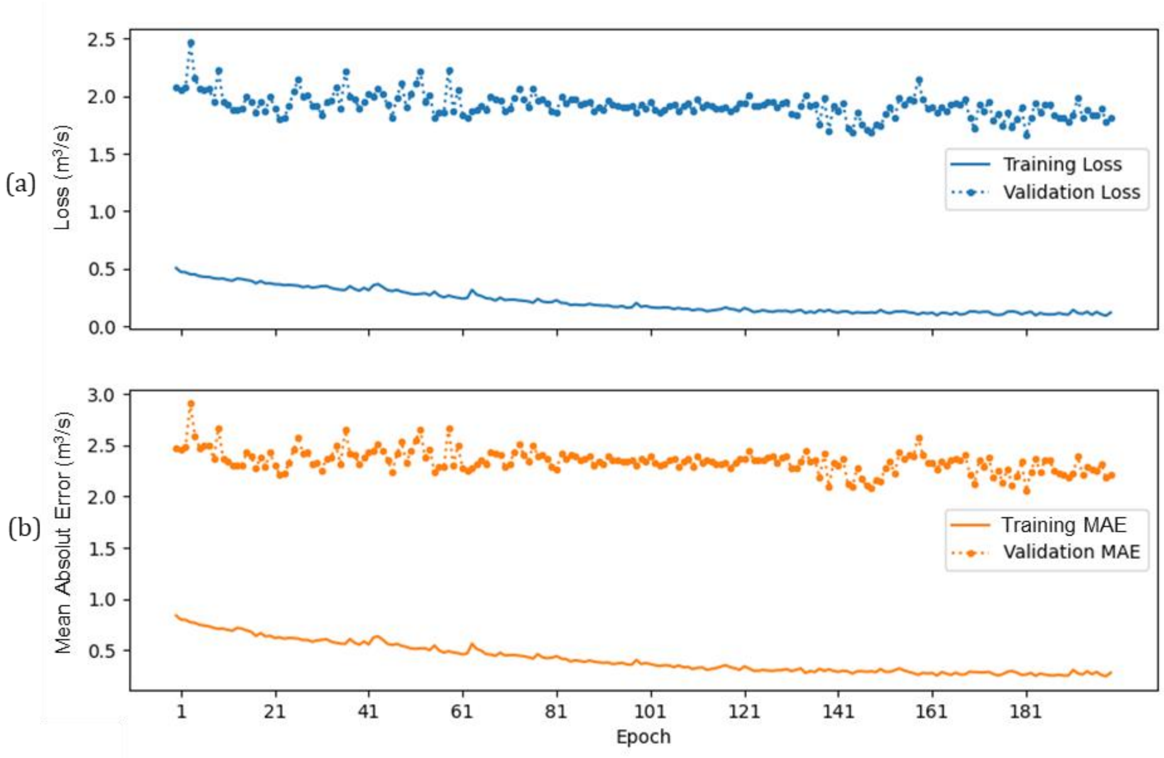


Figure 5 (a) Training and validation loss comparison over 200 epochs; (b) MAE and validation value trends.

The trained data was then evaluated on 100% of the dataset. A comprehensive evaluation of the model's predictive performance from 2006 to 2013 is shown on the hydrograph model in Figure 6a. This figure represents a full comparison between the model and the raw dataset. The hydrograph time series visualization in Figure 6b provides an overview of the model's performance on the test data, as well as comparing it to historical rainfall patterns. The hydrograph shows the model's ability to capture the overall flow trend and variability, with good agreement during periods of high and low flow. However, some discrepancies emerge during certain events, especially when there are rapid changes in streamflow values, indicating potential limitations in capturing extreme hydrological events.

Model accuracy was evaluated using the 1:1 line, which serves as a benchmark for comparing the fit of observed and modelled values (Figure 6c). Data clustering around the 1:1 line indicates a good fit between observed and modelled streamflow. Conversely, the degree of scatter, especially at higher flow ranges, indicates potential over- or under-prediction in some instances. Performance metrics show an RMSE of 0.74 and an NSE value of 0.44, which measures the model's ability to capture the variability and magnitude of the observed data. The computational efficiency of the model is indicated by the training time of 25253 seconds or approximately ±7 hours computational time.

Performance metrics for the testing data, as shown in Figure 6d, display significant deviation from the original datasets, exhibiting an RMSE of 1.57 and an NSE of −0.10. This suggests that the model may struggle to generalize its predictions to unseen data, emphasizing the necessity for further refinement or exploration of different computational model architectures.

The performance metrics further highlight the model's tendency to overpredict. Higher RMSE values on the partial test compared to the entire test dataset indicate a reduced ability of the model to capture the observed flow magnitudes over a given period. Similarly, significantly lower NSE values on the partial test highlight a reduced capacity of the model to reproduce the observed streamflow variability. These findings overall suggest that model performance is sensitive to data characteristics, with its predictive ability being compromised when applied to a subset of the original dataset.

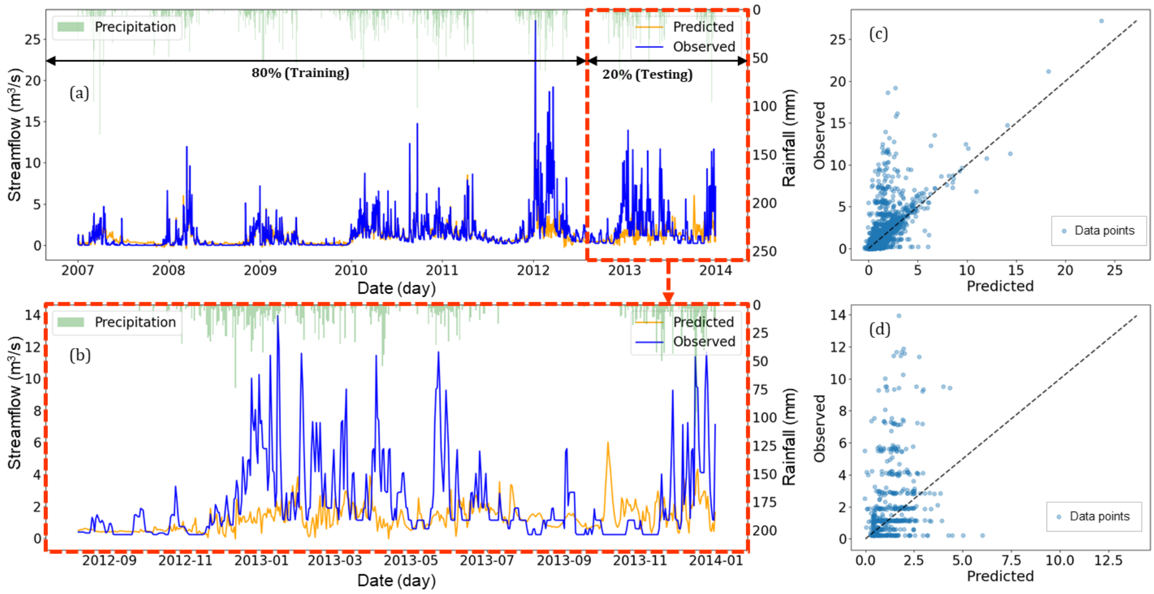


Figure 6 (a) Comparison of predicted vs observed river streamflow hydrographs; (b) model performance over the period of 2012–2013 (zoom in of Figure 6a); (c) Scatter plots confirm overprediction bias with most data points above 1:1 line; (d) 20% test dataset confirms consistent overestimation across various subsets.

Conclusion

This study presents the development and assessment of an LSTM model designed for predicting streamflow in the Pandanduri Reservoir Catchment Area in Indonesia. It involves a detailed examination of various hyperparameter settings, i.e., batch size, lag time, scaler, and loss function. An optimized model that features a batch size of 4, a lag time of 365 days, a robust scaler, and a Huber loss function achieves reasonable accuracy in simulating river flow across the full dataset, with an NSE of 0.44 and an RMSE of 0.74. The finding indicates that the model performance declines when applied on partial value, characterized by a decrease in NSE to -0.10 and an increase in RMSE to 1.57. This sensitivity to data characteristics suggests that, while the LSTM model can effectively learn catchment behavior from limited records, its generalization capability is constrained by how representative the training data are.

The hyperparameter analysis reveals several practical insights for future applications: robust scaling outperforms min-max normalization by preserving outlier information crucial for extreme event prediction; Huber loss provides better balance than MSE between fitting common events and accommodating extremes; and moderate batch sizes (16-32) with extended lag times (90-365 days) consistently outperform alternatives. These findings contribute to the growing body of knowledge on best practices for LSTM implementation in data-scarce tropical regions.

The future study should focus on addressing the limitations of the model, especially in predicting extreme hydrological events and improving performance on partial data. Key priorities are to: (1) add hydroclimate variables, such as groundwater levels and soil moisture indices, to better capture baseflow dynamics; (2) test ensemble methods that combine multiple LSTM architectures or hybrid LSTM–conceptual models to improve robustness; (3) use transfer learning from data-rich catchments in similar hydroclimatic regions to enhance extreme-event prediction; and (4) design physics-informed loss functions that encode hydrological constraints to ensure physically consistent outputs. A detailed analysis of how data characteristics influence model sensitivity is also needed to improve applicability across diverse hydrological regions and to encourage wider use in Indonesian reservoir management.

ACKNOWLEDGEMENTS

This study was financially supported by PPMI FTSL ITB 2025.

REFERENCES

Andika, N., Wongso, P., Rohmat, F. I. W., Wulandari, S., Fadhil, A., Rosi, R., & Burnama, N. S. (2025). Machine learning-based hydrograph modeling with LSTM: A case study in the Jatigede Reservoir Catchment, Indonesia. *Results in Earth Sciences*, *3*, 100090. https://doi.org/10.1016/j.rines.2025.100090

Aribowo, G. G., Juwono, P. T., Suhartanto, E., & Asmaranto, R. (2023). The Risk-Level Change of Dam Break Due to the Population Growth in the Dam Downstream. *Journal of Hunan University Natural Sciences*, *50*(1), 219–227. https://doi.org/10.55463/issn.1674-2974.50.1.22

BBWS. (2025). *Profil Bendungan Pandanduri*. https://sda.pu.go.id/balai/bbwsnt1/post/264/profil-bendungan-pandanduri

Biswas, N. K., Hossain, F., Biswas, N. K., & Hossain, F. (2022). A Multidecadal Analysis of Reservoir Storage Change in Developing Regions. *Journal of Hydrometeorology*, *23*(1), 71–85. https://doi.org/10.1175/JHM-D-21-0053.1

Black, A., Peskett, L., MacDonald, A., Young, A., Spray, C., Ball, T., Thomas, H., & Werritty, A. (2021). Natural flood management, lag time and catchment scale: Results from an empirical nested catchment study. *Journal of Flood Risk Management*, *14*(3), e12717. https://doi.org/10.1111/jfr3.12717

Burnama, N. S., Rohmat, F. I. W., Farid, M., Kuntoro, A. A., Kardhana, H., Rohmat, F. I. W., & Wijayasari, W. (2023). The Utilization of Satellite Data and Machine Learning for Predicting the Inundation Height in the Majalaya Watershed. *Water*, *15*(17), 3026. https://doi.org/10.3390/w15173026

de Moura, C. N., Seibert, J., & Detzel, D. H. M. (2022). Evaluating the long short-term memory (LSTM) network for discharge prediction under changing climate conditions. *Hydrology Research*, *53*(5), 657–667. https://doi.org/10.2166/nh.2022.044

Draper, D. W., Newell, D. A., Wentz, F. J., Krimchansky, S., & Skofronick-Jackson, G. M. (2015). The Global Precipitation Measurement (GPM) microwave imager (GMI): Instrument overview and early on-orbit performance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(7), 3452–3462. https://doi.org/10.1109/JSTARS.2015.2403303

Enung, Kusuma, M. S. B., Kardhana, H., Suryadi, Y., & Rohmat, F. I. W. (2022). Hourly Discharge Prediction Using Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) in The Upper Citarum River. *GEOMATE Journal*, *23*(98), 147–154. https://doi.org/10.21660/2022.98.3462

Hasan, F., Medley, P., Drake, J., & Chen, G. (2024). Advancing Hydrology through Machine Learning: Insights, Challenges, and Future Directions Using the CAMELS, Caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE Datasets. *Water*, *16*(13), 1904. https://doi.org/10.3390/w16131904

Hirsch, R. M., Moyer, D. L., & Archfield, S. A. (2010). Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs1. *JAWRA Journal of the American Water Resources Association*, *46*(5), 857–880. https://doi.org/10.1111/j.1752-1688.2010.00482.x

Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, *6*(4), 312–315. https://doi.org/10.1016/j.icte.2020.04.010

Kardhana, H., Valerian, J. R., Rohmat, F. I. W., & Kusuma, M. S. B. (2022). Improving Jakarta's Katulampa Barrage Extreme Water Level Prediction Using Satellite-Based Long Short-Term Memory (LSTM) Neural Networks. *Water*, *14*(9), 1469. https://doi.org/10.3390/w14091469

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018a). Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018b). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall-runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrology and Earth System Sciences*, *25*(10), 5517–5534. https://doi.org/10.5194/hess-25-5517-2021

Li, J., Qian, K., Liu, Y., Yan, W., Yang, X., Luo, G., & Ma, X. (2022). LSTM-Based Model for Predicting Inland River Runoff in Arid Region: A Case Study on Yarkant River, Northwest China. *Water*, *14*(11), 1745. https://doi.org/10.3390/w14111745

Majone, B., Avesani, D., Zulian, P., Fiori, A., & Bellin, A. (2022). Analysis of high streamflow extremes in climate change studies: how do we calibrate hydrological models? *Hydrology and Earth System Sciences*, *26*(14), 3863–3883. https://doi.org/10.5194/HESS-26-3863-2022

Maurya, V. (2024). Advancements and Challenges in Hydrological Modelling. *International Journal for Research in Applied Science and Engineering Technology*, *12*(11), 746–751. https://doi.org/10.22214/IJRASET.2024.65161

Michael, S., David, P., William, C., Martin, R. F., Sierks, M., Pierce, D. W., Chapman, W. E., & Ralph, F. M. (2022). Seasonally Anchored Bias Correction of CMIP5 Hydrological Simulations. *Authorea Preprints*. https://doi.org/10.1002/essoar.10512497.1

NASA. (2018). *Global Precipitation Measurement (GPM)*. NASA Global Precipitation Measurement (GPM). https://gpm.nasa.gov/data

Nguyen, N. Y., Kha, D. D., Van Ninh, L., Anh, V. T., & Anh, T. N. (2025). Streamflow prediction using Long Short-Term Memory networks: a case study at the Kratie Hydrological Station, Mekong River Basin. *Journal of Hydroinformatics*, *27*(2), 275–298. https://doi.org/10.2166/hydro.2025.276

Pathan, A. I., Sidek, L. B. M., Basri, H. B., Hassan, M. Y., Khebir, M. I. A. Bin, Omar, S. M. B. A., Khambali, M. H. bin M., Torres, A. M., & Najah Ahmed, A. (2024). Comparative assessment of rainfall-based water level prediction using machine learning (ML) techniques. *Ain Shams Engineering Journal*, *15*(7), 102854. https://doi.org/10.1016/j.asej.2024.102854

Sahu, R., Srivastava, S., & Kumar, R. (2023). Modelling of a Non-Linear Dynamic System Using Long Short-Term Memory. *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 1016–1021. https://doi.org/10.1109/ICCCIS60361.2023.10425027

Sharma, V. (2022). A Study on Data Scaling Methods for Machine Learning. *International Journal for Global Academic & Scientific Research*, *1*(1), 31–42. https://doi.org/10.55938/ijgasr.v1i1.4

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Snider, E., Iseri, Y., Trinh, T., Anderson, M. L., & Kavvas, M. L. (2024). Reconstructing 169 years of historical atmospheric and hydrologic conditions in the American River Watershed through the Watershed Environmental Hydrology Hydro-Climate Model system. *Frontiers in Water*, *6*, 1445722. https://doi.org/10.3389/FRWA.2024.1445722/TEXT

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Sushanth, K., Mishra, A., & Singh, R. (2023). Real-time reservoir operation using inflow and irrigation demand forecasts in a reservoir-regulated river basin. *Science of The Total Environment*, *904*, 166806. https://doi.org/10.1016/j.scitotenv.2023.166806

Taye, M. T., Zimale, F. A., Woldesenbet, T. A., Kebede, M. G., Amare, S. D., Tegegne, G., Mekonnen, K., & Haile, A. T. (2023). Priority Research Topics to Improve Streamflow Data Availability in Data-Scarce Countries: The Case for Ethiopia. *Hydrology 2023, Vol. 10*, *10*(12). https://doi.org/10.3390/hydrology10120220

Trimartinni, A. D., Limantara, L. M., Juwono, P. T., & Siswoyo, H. (2024). Limiting Parameters for Supporting Performance Index Assessment of Multi-Purpose Reservoir Storage Capacity. *Journal of Hunan University Natural Sciences*, *51*(6). https://doi.org/10.55463/issn.1674-2974.51.6.24

Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2022). A Comprehensive Survey of Loss Functions in Machine Learning. *Annals of Data Science*, *9*(2), 187–212. https://doi.org/10.1007/s40745-020-00253-5

Xiang, K., Li, Y., Horton, R., & Feng, H. (2020). Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. *Agricultural Water Management*, *232*, 106043. https://doi.org/10.1016/j.agwat.2020.106043

Xiang, Z., Yan, J., & Demir, I. (2020). A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. *Water Resources Research*, *56*(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

Yadav, D., Sahoo, L., Mandal, S. K., Ravivarman, G., Vijayaraghavan, P., & Prasad, B. (2023). Using Long Short-Term Memory Units for Time Series Forecasting. *2023 2nd International Conference on Futuristic Technologies, INCOFT*

2023. https://doi.org/10.1109/INCOFT60753.2023.10425756

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, *598*, 126266. https://doi.org/10.1016/j.jhydrol.2021.126266