Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Using Machine Learning and Satellite Data for Improving Flood Prediction in the Lake Tanganyika Basin: A Case Study of the Rusizi Sub-Basin

Jean claude Hakizimana

alvinhakizimana@gmail.com

Université d'Abomey-Calavi https://orcid.org/0009-0004-8490-6914

N'Tcha M'Po Yèkambèssoun

Université d'Abomey-Calavi

Emmanuel Egnide Lawin

Université d'Abomey-Calavi

Freddy Hounoundaho

National University of Science, Technology, Engineering and Mathematics, Abomey-Calavi, Benin

Research Article

Keywords: Flood forecasting, ConvLSTM, LSTM, ERA5, Artificial Intelligence, Rusizi Basin, Satellite data

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11753750/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Using Machine Learning and Satellite Data for Improving Flood Prediction in the Lake Tanganyika Basin: A Case Study of the Rusizi Sub-Basin

Jean Claude Hakizimana[1,2], N'Tcha M'Po Yèkambèssoun[2], Emmanuel Egnide Lawin[2], and Freddy Hounoundaho[2,3]

[1] Climarys, Bujumbura, Burundi
[2] Laboratory of Applied Hydrology, National Institute of Water, University of Abomey-Calavi, Benin
[3] National University of Science, Technology, Engineering and Mathematics, Abomey-Calavi, Benin
alvinhakizimana@gmail.com

Abstract

Floods are among the most destructive natural hazards, particularly in Africa, where climate change, rapid urbanization, and data scarcity amplify their impacts. The Rusizi sub-basin has experienced recurrent floods that have caused significant disruptions and socioeconomic effects from 2014 to 2024. This study investigates the potential of machine learning, specifically Long Short-Term Memory (LSTM) and Convolutional LSTM (ConvLSTM) models, combined with satellite and hydrological datasets, to enhance flood prediction in the Rusizi Basin. Using ERA5 reanalysis meteorological data, river discharge , and satellite derived DEMs, the ConvLSTM model achieved high predictive skill ($R^2 = 0.96$, NSE $= 0.96$, RMSE ≈ 0.0023), outperforming standard LSTM ($R^2 = 0.96$). Flood risk maps were also generated for water levels of 4 m, 10 m, and 12 m. Results demonstrate that integrating AI-based models with satellite data significantly improves flood forecasting, providing valuable decision support tools for early warning and disaster risk reduction. Future integration with real-time monitoring and socioeconomic vulnerability data could further strengthen climate resilience strategies.

Keywords: Flood forecasting, ConvLSTM, LSTM, ERA5, Artificial Intelligence, Rusizi Basin, Satellite data

1 Introduction

Floods are among the most frequent and destructive natural hazards worldwide [6], accounting for nearly 74% of water-related disasters between 2001 and 2018 [27]. Their impacts extend beyond physical damage, threatening lives, livelihoods, and infrastructure, with recovery often prolonged in low- and middle-income countries, reinforcing poverty and vulnerability [2]. In Africa, flood risk is intensifying under climate change, rapid urbanization, and population growth in flood-prone areas [1]. The Rusizi Basin exemplifies these challenges. Playing a critical role in regional food security, hydropower, and trade, the basin's communities remain highly exposed to recurrent floods. Between 2014 and 2023, repeated events displaced populations, destroyed farmland, and disrupted livelihoods, underscoring the basin's socio-economic vulnerability. Managing such hazards is complicated by fragmented monitoring networks, limited data-sharing, and differing institutional capacities among riparian states.

Traditional hydrological models, reliant on rainfall-runoff relationships or physical hydraulics, provide detailed insights when supported by dense observations. However, their effectiveness is severely limited in data-scarce regions where gauging records, meteorological stations, and topographic data are sparse [2]. The nonlinear nature of flood processes in tropical catchments further challenges model calibration and transferability. Recent advances in artificial intelligence (AI) and Earth observation (EO) offer new opportunities to address these limitations. Reanalysis products like ERA5 [7] provide long-term, spatially continuous meteorological data, while global DEMs such as SRTM [24] improve terrain accessibility. Combined with satellite-based hydrological observations, these resources reduce dependence on ground measurements, enabling flood forecasting in data-constrained regions.

Machine learning approaches have emerged as powerful tools for modeling nonlinear spatio-temporal processes. Long Short-Term Memory (LSTM) networks [8] capture temporal dependencies, while Convolutional LSTM (ConvLSTM) models [21] incorporate spatial context. Neural operators like Fourier Neural Operators (FNO) further enable surrogate modeling of partial differential equations [26]. Collectively, these advances signal a paradigm shift in flood prediction research. Despite rapid progress, AI-based hydrological modeling remains limited in African transboundary basins, with most studies focused on data-rich environments. Critical gaps persist where observational networks and institutional capacity are weak. Moreover, machine learning models have yet to be fully integrated into operational early warning systems, and questions remain about handling uncertainty, incorporating physical understanding, and informing decision-making in vulnerable communities. This study addresses these gaps by applying modern machine learning in a data-scarce, transboundary African basin. Using the Rusizi Basin as a test case, we evaluate advanced AI models' capacity to capture flood dynamics under limited *in situ* observations but growing satellite and reanalysis data availability, contributing to discussions on AI's role in enhancing disaster preparedness, climate resilience, and sustainable development in at-risk regions.

2 Study Area and Data

2.1 Study Area

The Rusizi Basin is a transboundary catchment in the Great Lakes region of Central Africa, lying between 2°30′S – 4°30′S latitude and 28°30′E – 29°30′E longitude, covering approximately 10,475 km^2 across Burundi, Rwanda, and the Democratic Republic of Congo (DRC), with the Rusizi River flowing southwards from Lake Kivu into Lake Tanganyika (Fig. 1). The basin experiences a tropical climate with a wet season from September to May, during which heavy rainfall frequently triggers flooding, and a dry season from June to August with reduced precipitation and relatively stable hydrological conditions. Socio-economic activities rely heavily on agriculture, fishing, and small-scale trade, yet dense population, rapid urbanization, and limited infrastructure increase vulnerability to flooding, with agricultural lands, homes, and transportation networks often severely affected during major events.

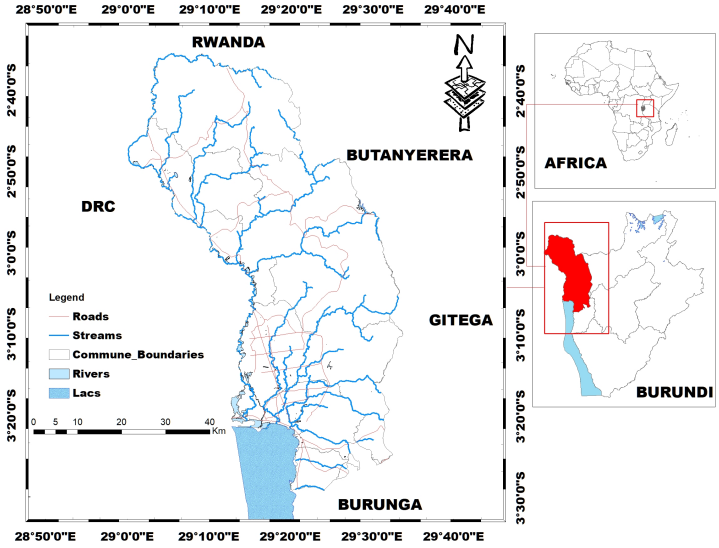


Fig. 1: Study area

2.2 Data Sources

This study employs multiple datasets to capture the meteorological, hydrological, and topographic factors controlling floods in the Rusizi Basin (Table 1). Meteorological information was derived from ERA5 and ERA5-Land reanalysis datasets (ECMWF) [7,25], including precipitation, air temperature, relative humidity, and wind speed spanning 2001–2020, all aggregated to daily resolution for consistency with hydrological observations. Hydrological data consist of daily river discharge measurements collected at IGEBU gauging stations over the same period. Topographic and satellite datasets include the SRTM Digital Elevation Model (USGS) [28] at 30 m resolution, complemented by Sentinel-2 optical imagery (ESA) [4] at 10–20 m resolution for 2015–2020.

Table 1: Summary of datasets used in the study

Dataset	Source	Resolution	Period
ERA5 Reanalysis (precipitation, air temperature, relative humidity, wind speed)	ECMWF	0.25°, hourly	2001–2020
River discharge(Rusizi River)	IGEBU	Point, daily	2001–2020
SRTM DEM	USGS	30 m	Static
Sentinel-2 Optical	ESA	10–20 m	2015–2020

3 Methodology

3.1 Problem Framing

Flood forecasting in the Rusizi Basin is formulated as a supervised learning problem following [13,20]. Let $\mathbf{x}_t \in \mathbb{R}^m$ denote the meteorological predictors at time t (precipitation, air temperature, relative humidity, and wind speed

from ERA5), and $Q_t \in \mathbb{R}$ represent the observed river discharge (m^3/s) at IGEBU gauging stations, which serves as the primary flood risk indicator and target variable. The input sequence of length k is defined as $\mathbf{X}_t^{(k)} = \{\mathbf{x}_{t-k+1}, \ldots, \mathbf{x}_t\}$, and the prediction model f_θ with parameters θ is expressed as $\hat{Q}_{t+\tau} = f_\theta(\mathbf{X}_t^{(k)})$, where $\hat{Q}_{t+\tau}$ denotes the predicted discharge at forecast horizon τ.

The dataset was split chronologically to preserve temporal dependencies: 70% training (2001–2014), 15% validation (2015–2017), and 15% testing (2018–2020). Input sequence length was set to $k = 24$ hours based on the basin's hydrological response time.

3.2 Model Development

Two deep learning architectures were investigated: Long Short-Term Memory (LSTM) and Convolutional LSTM (ConvLSTM).

Long Short-Term Memory (LSTM) The LSTM is a recurrent neural network designed to capture temporal dependencies in sequential data through gating mechanisms. The mathematical formulation is expressed as:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \tag{1}$$
$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \tag{2}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \tag{3}$$
$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \tag{4}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{5}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{6}$$

where $\mathbf{x}_t$ is the input vector at time t, $\mathbf{h}_t$ is the hidden state (short-term memory), $\mathbf{c}_t$ is the cell state (long-term memory), $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ are the input, forget, and output gates controlling information flow, $\tilde{\mathbf{c}}_t$ is the candidate cell state, $\mathbf{W}.$ and $\mathbf{U}.$ are weight matrices, $\mathbf{b}.$ are bias vectors, σ is the sigmoid activation function, tanh is the hyperbolic tangent, and $\odot$ denotes element-wise multiplication.

Convolutional LSTM (ConvLSTM) ConvLSTM extends LSTM by replacing matrix multiplications with convolution operations to capture both spatial and temporal dependencies:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i * \mathbf{X}_t + \mathbf{U}_i * \mathbf{H}_{t-1} + \mathbf{b}_i) \tag{7}$$
$$\mathbf{f}_t = \sigma(\mathbf{W}_f * \mathbf{X}_t + \mathbf{U}_f * \mathbf{H}_{t-1} + \mathbf{b}_f) \tag{8}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o * \mathbf{X}_t + \mathbf{U}_o * \mathbf{H}_{t-1} + \mathbf{b}_o) \tag{9}$$
$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_c * \mathbf{X}_t + \mathbf{U}_c * \mathbf{H}_{t-1} + \mathbf{b}_c) \tag{10}$$
$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \tag{11}$$
$$\mathbf{H}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \tag{12}$$

where $*$ denotes the 2D convolution operation, $\mathbf{X}_t$ is a tensor representing spatial inputs (e.g., a precipitation map), $\mathbf{H}_{t-1}$ and $\mathbf{H}_t$ are 3D tensors representing the hidden state with spatial dimension, $\mathbf{C}_{t-1}$ and $\mathbf{C}_t$ are 3D cell state tensors, and $\mathbf{W}.$ and $\mathbf{U}.$ are convolution kernels. This architecture captures both spatial dependencies through convolutions and temporal dependencies through recurrence, making it particularly suitable for spatio-temporal flood dynamics.

3.3 Hyperparameter Optimization and Training

Model hyperparameters such as learning rate (η), number of ConvLSTM filters (U), dropout rate (d), epochs (E), and batch size (B) were optimized using Bayesian optimization with a Gaussian Process surrogate (scikit-optimize). The search space is shown in Table 2. Starting from default values ($\eta = 0.01$, $U = 100$, $d = 0.1$, $E = 100$, $B = 100$), the optimization ran for 11 iterations using Expected Improvement acquisition, selecting the best parameters based on validation loss. Early stopping (patience=5) prevented overfitting. Performance was evaluated using RMSE, NSE, and R^2 metrics:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2} \tag{13}$$
$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{N}(Q_i - \bar{Q})^2} \tag{14}$$
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{N}(Q_i - \bar{Q})^2} \tag{15}$$

where Q_i is observed discharge, $\hat{Q}_i$ predicted discharge, $\bar{Q}$ mean observed discharge, and N number of observations.

Table 2: Hyperparameter search space

Parameter	Symbol	Range	Prior
Learning rate	η	$[10^{-3}, 1]$	Log-uniform
ConvLSTM filters	U	$[1, 300]$	Uniform
Dropout rate	d	$[0.0, 0.5]$	Uniform
Epochs	E	$[1, 300]$	Uniform
Batch size	B	$[1, 100]$	Uniform

3.4 Flood Risk Mapping

The LSTM output models were used to predict river discharge (m^3/s) for various flood events. These discharge predictions were converted to corresponding water levels (in meters) using a rating curve (stage-discharge relationship) established from historical measurements at IGEBU gauging stations. The predicted water levels, ranging from 1 m to 12 m, were then integrated into Global Mapper, a GIS-based hydraulic modeling software, along with SRTM-derived elevation data (30 m resolution). This integration enabled the simulation of 3D inundation extents for each water level scenario.

The resulting flood hazard maps illustrate the progressive expansion of inundation from 1 m to 12 m water levels, highlighting critical infrastructure and priority zones for flood risk management. Three representative scenarios4 m, 10 m, and 12 m were selected to demonstrate low, medium, and high flood severity levels, respectively, providing actionable information for disaster preparedness and early warning. Figure 4 presents these representative flood hazard maps for the Rusizi Basin.

To couple hydrological forecasts with risk assessment, flood risk $R(x,t)$ is defined as the product of hazard, exposure, and vulnerability:

$$R(x,t) = H(x,t) \cdot E(x,t) \cdot V(x,t) \tag{16}$$

where $H(x,t)$ represents water depth or inundation extent, $E(x,t)$ captures the presence of populations, infrastructure, and economic activities, and $V(x,t)$ measures the susceptibility of assets to the hazard.

3.5 System Architecture

The complete workflow (Fig. 2) integrates data acquisition (ERA5, USGS, GloFAS, satellite), preprocessing (cleaning, normalization, sequence generation), model training, performance evaluation, and flood risk mapping with decision support.

4 Results

4.1 Characteristics of Meteorological and Hydrological Variables

The statistical characteristics of the meteorological and hydrological variables used for model development are summarized in Table 3. Temperature showed moderate variability, while precipitation and discharge exhibited strong skewness associated with extreme rainfall and flood events. It highlights the nonlinear nature of the hydrological response in the basin and justifies the use of deep learning approaches.

Table 3: Descriptive statistics of meteorological and hydrological variables

Variable	Mean	Std.	Min	25%	50%	75%	Max
Temperature (°C)	26.93	4.26	14.77	23.45	26.19	30.02	38.91
Wind u10 (m/s)	0.054	1.15	-4.18	-0.74	0.24	0.87	4.87
Wind v10 (m/s)	0.67	1.50	-3.75	-0.26	0.96	1.79	5.43
Precipitation (mm)	0.14	0.43	0.00	0.00018	0.0017	0.04	10.07
Pressure (Pa)	97041.78	219.30	96111.91	96896.32	97054.98	97199.23	97743.73
Discharge (m^3/s)	54.71	100.70	0.00	0.30	4.35	54.71	621

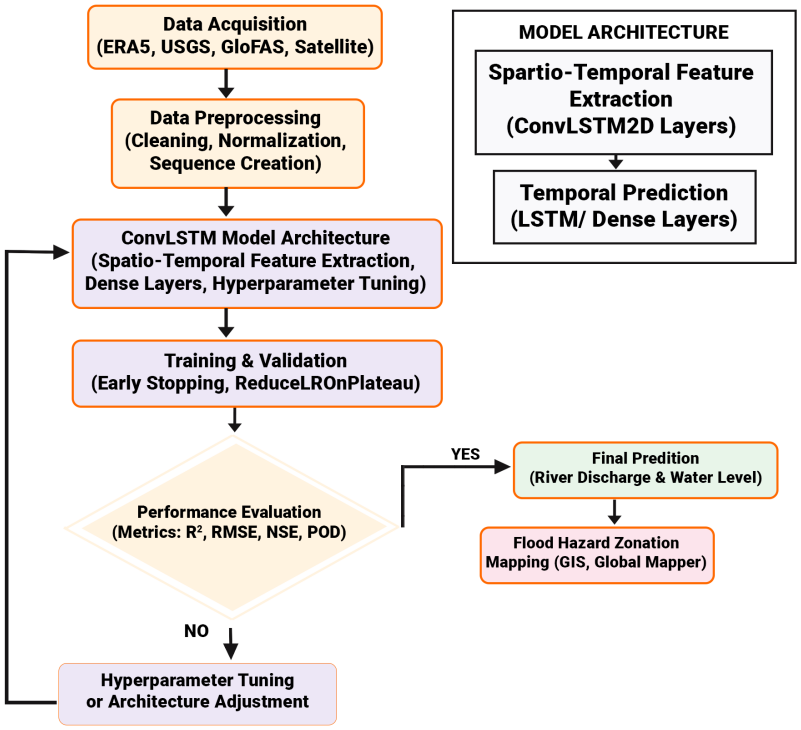


Fig. 2: Flood Prediction System Architecture with ConvLSTM Model

4.2 Model Performance

Both models demonstrated strong predictive capability for forecasting river discharge in the Rusizi Basin, with the ConvLSTM slightly outperforming the standard LSTM. The ConvLSTM achieved an R^2 of 0.95, NSE of 0.95, and RMSE of 0.0023, while the LSTM reached $R^2 = 0.95$, NSE = 0.96, and RMSE = 0.046 (Table 4). The skill of the ConvLSTM reflects its ability to capture abrupt flow fluctuations and peak discharges, benefiting from the convolutional layers that enhance spatio-temporal learning. In contrast, the LSTM reproduced the general discharge dynamics well but tended to underestimate extreme events. These results are consistent with recent regional [3] and global studies [17,15], confirming the added value of hybrid architectures for hydrological forecasting.

Table 4: Performance metrics of ConvLSTM and LSTM models

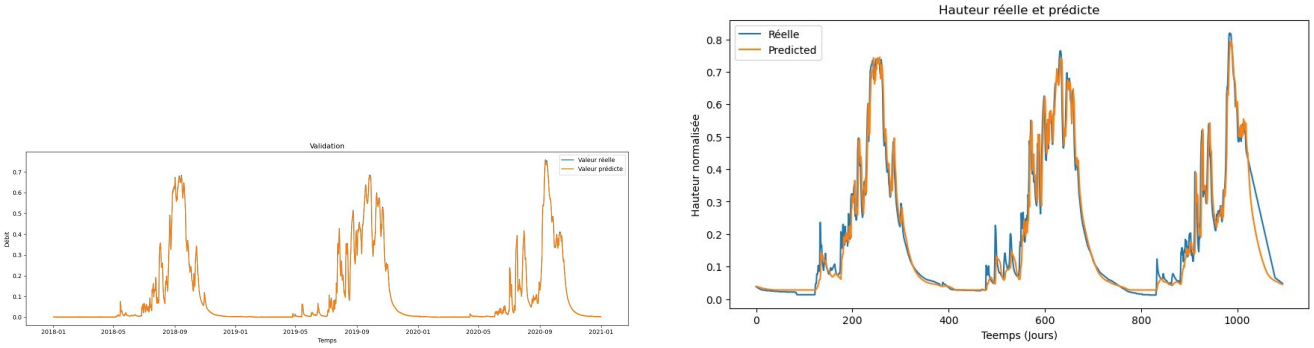
Model	R^2	NSE	RMSE
ConvLSTM	0.95	0.95	0.0023
LSTM	0.95	0.96	0.046

4.3 Model Validation

Validation on an independent test dataset confirmed the robustness of both deep learning models (Fig. 3). The ConvLSTM accurately reproduced the observed discharge hydrograph, including the sharp peaks associated with extreme rainfall events. In contrast, the LSTM achieved good overall agreement but tended to underestimate the highest discharge peaks, indicating a lower capability for capturing extreme flood conditions.

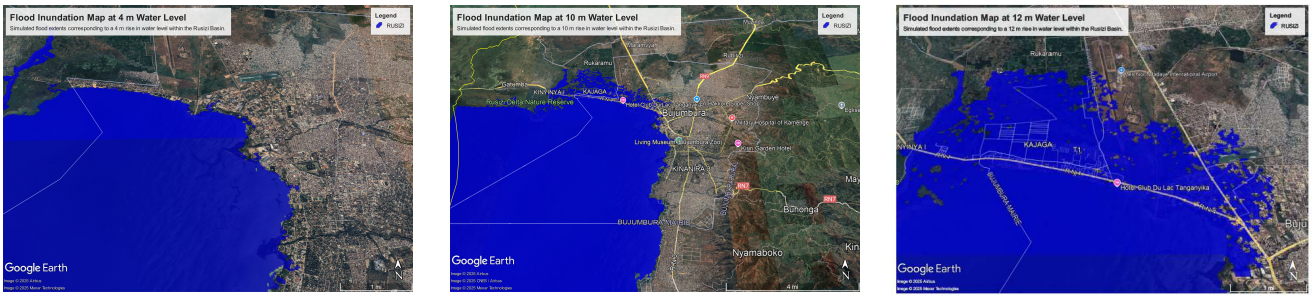
Following validation, the ConvLSTM model was used to generate flood hazard maps for three representative water-level scenarios (4 m, 10 m, and 12 m), illustrating the progression of inundation across the Rusizi Basin (Fig. 4). At a water level of 4 m, flooding remains largely confined to riverbanks and adjacent agricultural land. At 10 m, inundation expands into residential areas, transportation networks, and critical infrastructure. Under the extreme 12 m scenario, widespread flooding occurs, resulting in substantial impacts on settlements and regional connectivity.

These flood hazard maps provide valuable support for flood risk management by identifying areas exposed under different flood scenarios. They can assist local authorities in emergency preparedness, evacuation planning, land-use management, and the implementation of effective early warning strategies.



(a) ConvLSTM (b) LSTM

Fig. 3: Validation of the ConvLSTM and LSTM models against observed river discharge using the independent test dataset.



(a) 4 m water level (≈3.74 ha inundated) (b) 10 m water level (≈6.31 ha inundated) (c) 12 m water level (≈7.17 ha inundated)

Fig. 4: Flood hazard maps for three representative water-level scenarios (4 m, 10 m, and 12 m). The corresponding inundated areas were estimated using Eq. (17).

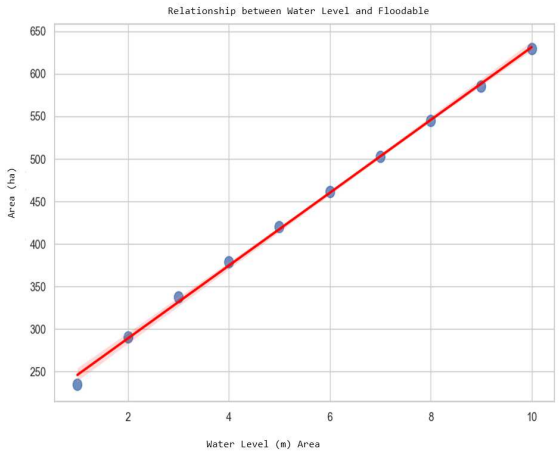


Fig. 5: Relationship between water level and inundated area used for flood extent estimation.

Figure 5 shows that the relationship between water level and inundated area can be approximated by the following linear regression model:

$$y = 0.4291x + 2.0231, \tag{17}$$

where y represents the inundated area (ha) and x denotes the water level (m). The regression achieved an excellent fit ($R^2 = 0.9986$), indicating a nearly linear relationship between water level and inundated area. The standard error of the regression slope was 0.5589, reflecting the uncertainty associated with the estimated coefficient.

4.4 Decision Support Indicators

Beyond flood hazard mapping, quantitative decision-support indicators were derived to strengthen flood early warning and disaster risk management. Flood risk was estimated using Eq. (**??**), with the associated variables

defined in Sect. 3.4. Additional indicators, including the probability of threshold exceedance ($p_{t+\tau}$) and the expected damage ($\mathbb{E}[D]$), provide forward-looking information to support preparedness planning, resource allocation, and anticipatory actions before flood events.

5 Discussion

The findings demonstrate the transformative potential of AI-based models for hydrological forecasting in data-scarce regions such as the Rusizi Basin. By leveraging satellite and reanalysis products, ConvLSTM and LSTM models help overcome the limitations of traditional lumped hydrological approaches, which have historically struggled in Sub-Saharan Africa due to sparse *in situ* observations [10,13]. The added capability of convolutional layers allows ConvLSTM to capture spatial rainfall–runoff dynamics more effectively, positioning it as a promising tool for regions where ground measurements are limited. These results align with previous studies showing that machine learning architectures, when combined with Earth observation datasets, can rival or even surpass process-based models in ungauged or poorly gauged basins [14,9].

In similar tropical and data-poor environments, such as the Congo Basin and neighboring catchments in Central Africa, hybrid AI methods have also demonstrated strong predictive skill. For example, machine learning and remote sensing combinations have been successfully used to model streamflow in large Central African basins, highlighting their capacity to capture complex hydrological responses where traditional monitoring networks are sparse [12,16]. Related research in Cameroon and Gabon further confirms that deep learning frameworks outperform empirical models in reproducing seasonal and extreme flow variability in tropical basins [5].

Beyond predictive accuracy, the potential operational value of such frameworks lies in their integration with ensemble forecasting, socio-economic exposure datasets, and transfer learning, which collectively could enhance scalability and adaptability across basins [11,22,19]. In Central African contexts, prototype AI-enabled forecasting systems have begun to demonstrate added utility when paired with inundation mapping and early warning platforms, pointing to broader regional relevance [23]. Moreover, addressing uncertainty through ensemble methods and transparent benchmarking ensures reproducibility and strengthens user trust in real-world applications [2,18].

From a risk management perspective, embedding these AI-driven forecasts into early warning systems could enable timely evacuation alerts, risk-informed decision-making, and prioritization of critical infrastructure protection, while coupling with community preparedness could foster long-term resilience. This is especially pertinent in transboundary basins like the Rusizi and Congo, where coordination among riparian states and data sharing are essential for effective hazard governance [1,27].

Beyond the technical results, this study opens important operational perspectives. Although the actual integration into local and regional early warning systems has not been realized within the scope of this work, the developed methodological framework could eventually be embedded into existing platforms. In the Rusizi Basin, collaboration with the National Hydrometeorological Service (IGEBU), civil protection agencies, and local communities would be essential to translate forecasts into actionable alerts. The proposed ConvLSTM-based framework could complement tools such as GloFAS or FEWS NET by offering higher spatial resolution and locally calibrated predictions, thus addressing the gap between global products and local needs. Operationalizing these forecasts within national disaster risk management strategies, coupled with community engagement, would constitute a major advancement for flood resilience.

6 Conclusion

This study has demonstrated the considerable potential of integrating satellite-derived datasets with advanced machine learning techniques to enhance flood forecasting in data-scarce river basins, exemplified by the Rusizi sub-basin. By combining ERA5 reanalysis data, SRTM-based topography, and local river discharge records, both LSTM and ConvLSTM models effectively captured the temporal and spatial dynamics of river flows, with ConvLSTM proving superior in detecting abrupt peaks and localized flood events—highlighting the added value of incorporating spatial convolution layers into temporal forecasting frameworks. The outcomes underscore the ability of AI-driven models to overcome challenges associated with limited *in situ* observations, providing timely and reliable flood predictions. The resulting flood hazard maps, generated from model outputs and high-resolution topography, serve as practical tools for decision-makers, enabling better planning, early warning, and preparedness. Moreover, the methodology demonstrates a scalable approach adaptable to other basins within the region, helping to bridge the gap between global forecasting products and localized risk management needs.

While this work presents a robust proof of concept, several limitations must be acknowledged. The scarcity of high-resolution local observations constrains fine-scale hydrological process representation, and the model's transferability to other catchments with different topographic and hydrological conditions remains untested. Additionally, the current framework focuses primarily on physical flood hazard without incorporating socio-economic vulnerability, infrastructure exposure, or adaptive capacity, limiting the comprehensiveness of risk assessments. Future work should therefore prioritize operationalizing the forecasting system into real-time early warning platforms, enriching the modeling framework with additional in situ measurements and higher-resolution remote sensing data,

assessing model transferability across other sub-basins, integrating socio-economic data for more comprehensive risk evaluation, and fostering community engagement with local authorities and disaster management agencies. By pursuing these directions, the framework can evolve from a research prototype into an operational decision-support system that contributes meaningfully to flood preparedness, climate resilience, and sustainable development in flood-prone regions.

References

1. Anyah, R.O., Semazzi, F.H.M.: Hydrological modeling in east and central african transboundary basins: challenges and opportunities. Water Resources Research **42** (2006). https://doi.org/10.1029/2006WR005059
2. Beven, K., Cloke, H.: Uncertainty in Hydrological Modelling and Forecasting. Cambridge University Press (2020)
3. Demetrakopoulos, I.: Short-term precipitation forecasting in the netherlands using convlstm. Journal of Hydrometeorology (2023)
4. ESA: Sentinel-1 and sentinel-2 mission documentation (2020)
5. Fotsing, M., Nyassé, S., Tchapda, A.: Deep learning versus conceptual hydrological models for streamflow forecasting in tropical river basins. Water Resources Management **37**, 3125–3142 (2023). https://doi.org/10.1007/s11269-023-03422-z
6. Hakizimana, J.C.: Développement d'un Modèle Simplifié de Prédiction des Précipitations au Burundi. Master's thesis, Université d'Abomey Calavi (Apr 2024). https://doi.org/10.13140/RG.2.2.18370.59841, advisor: Prof. Agnidé Emmanuel Lawin, Dr. Yékambessoun N'Tcha M'PO
7. Hersbach, H., et al.: The era5 global reanalysis. Quarterly Journal of the Royal Meteorological Society **146**(730), 1999–2049 (2020)
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
9. Hountondji, Y.C., Diedhiou, A., Mbow, C.A.: Use of era5 and machine learning approaches for rainfall–runoff modelling in west and central africa. Hydrology Research **51**(6), 1550–1568 (2020). https://doi.org/10.2166/nh.2020.137
10. Kauffeldt, A., et al.: A framework for benchmarking large-scale hydrological models. Water Resources Research (2016)
11. Koks, E.E., et al.: Integrated flood risk modelling: Combining exposure, vulnerability and hazard. Nature Climate Change (2019)
12. Koukpaï, A., Tchameni, R., Nkongho, R.: Modeling streamflow in the congo river basin using machine learning and remote sensing data. Journal of Hydrology: Regional Studies **22**, 100597 (2019). https://doi.org/10.1016/j.ejrh.2019.100597
13. Kratzert, F., et al.: Towards learning universal, regional, and local hydrological models using lstms. Hydrology and Earth System Sciences (2019)
14. Kumar, R., Ahmed, S., Jain, S.K.: Machine learning in hydrological applications: A review. Environmental Modelling & Software **126**, 104654 (2020). https://doi.org/10.1016/j.envsoft.2020.104654
15. Liu, B., et al.: Machine learning in flood depth estimation: A review of 108 studies. SpringerLink (2025)
16. Mbow, C.A., Diedhiou, A., Sene, E., Hountondji, Y.C.: Data-driven river discharge predictions in ungauged sub-basins of central africa. Environmental Modelling & Software **137**, 104957 (2021). https://doi.org/10.1016/j.envsoft.2021.104957
17. Mosavi, A., et al.: Flood prediction using machine learning models: A review. Sustainability (2018)
18. Nerini, D., et al.: Uncertainty quantification for machine learning hydrology. Water Resources Research (2022)
19. Pappenberger, F., et al.: Operational flood forecasting and ensemble prediction. Hydrology and Earth System Sciences (2020)
20. Shen, C., et al.: Hybrid modeling for streamflow forecasting: Integrating physics and machine learning. Water Resources Research (2020)
21. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.k., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 28 (2015)
22. Shrestha, D., et al.: Transfer learning for streamflow prediction in data-scarce basins. Hydrology and Earth System Sciences (2021)
23. Simo, J., Fokou, G., Kenfack, A.: Ai-enabled flood forecasting framework for the congo basin: potentials and challenges. Natural Hazards **114**, 237–259 (2022). https://doi.org/10.1007/s11069-022-05312-1
24. Smith, M.J., et al.: On the use of hydrodynamic models for flood inundation mapping. Hydrology and Earth System Sciences **19**(10), 4325–4338 (2015)
25. Store, C.C.D.: Era5-land documentation (2020), available at: https://cds.climate.copernicus.eu
26. Sun, A., Li, Z., Lee, W., Huang, Q., Scanlon, B., Dawson, C.: Rapid flood inundation forecast using fourier neural operator. arXiv preprint arXiv:2303.00000 (2023)
27. United Nations Office for Disaster Risk Reduction (UNDRR): Global assessment report on disaster risk reduction (2019)
28. USGS: Shuttle radar topography mission (srtm) dem (2020), digital Elevation Data