A Deep Learning Approach to Flood Prediction and Early Warning Using Multi-Source Environmental Data: Evidence from Zimbabwe

Tavengwa Norman[1] and Fungai Hwacha[2]

[1]School of Information Science and Technology, Harare Institute of Technology, Harare, Zimbabwe

[2]Environmental Science, University of Indonesia, Jakarta, Indonesia

DOI: https://doi.org/10.51244/IJRSI.2026.1304000089

Received: 30 March 2026; Accepted: 04 April 2026; Published: 02 May 2026

ABSTRACT

Zimbabwe ranks among the most flood-prone countries in Southern Africa; however, its existing flood warning systems lack adequate coverage and depend on insufficiently advanced technologies. In this study, we propose an applicable deep learning-based framework combining several sources of satellite, topographic, and in-situ data that would facilitate flood predictions and early warning systems implementation in three Zimbabwean catchments, the Save, Manyame, and Mazowe. Seven models were created and evaluated in this study; one of them, hybrid CNN-LSTM model, demonstrated better performance results with 95.9% accuracy, F1-score of 95.0%, and AUC-ROC of 0.981 for the independent test set. In addition, spatial cross-validation was applied to prove the generalization capacity of the proposed model. According to SHAP analysis, the following predictors were determined as the most influential: antecedent rainfall within 72 hours, distance to the closest river channel, and Terrain Wetness Index, all of which coincide with real-life features of Zimbabwe's hydrology. As for the end user, the suggested model could be incorporated into a four-level flood early warning system (advisory level, watch level, warning level, and emergency level).

Keywords: CNN-LSTM, Flood prediction, Early warning systems, multi-source data fusion, SHAP

INTRODUCTION

Zimbabwe is considered one of the most climate-vulnerable countries in Sub-Saharan Africa. Its hydroclimatic conditions are marked by significant fluctuations in rainfall from year to year, recurrent droughts, and occasional, yet extremely damaging, floods [1]. Flooding constitutes the most widespread and commonly occurring natural hazard within the nation, responsible for most disaster-related fatalities, population displacements, and infrastructure damage documented from 1980 to 2023 [2]. The destructive impact of Tropical Cyclone Idai, which struck the Chimanimani and Chipinge districts in March 2019, resulting in 344 confirmed deaths, the displacement of over 270,000 individuals, and infrastructure damage estimated at roughly USD 622 million, exemplifies the most catastrophic flood event in Zimbabwe's recent history and highlights the urgent need for enhanced early warning systems [3].

Beyond cyclone-driven extremes, seasonal river flooding along the Save, Manyame, Mazowe, Runde, and Mzingwane systems regularly destroys smallholder agricultural production, contaminates rural water supplies, and cuts transport links across large portions of the country's rural hinterland [4].Operational flood forecasting in Zimbabwe is the primary responsibility of the Zimbabwe National Water Authority (ZINWA) and the Meteorological Services Department (MSD), which maintain networks of streamflow gauging stations and surface weather observation posts respectively. Conversely, Zimbabwe's hydrological monitoring infrastructure has suffered from consistent decline since the mid-2000s, a consequence of inadequate funding, instrument damage, and the destabilizing impact of hyperinflation on maintenance allocations [5]. By 2023, less than 40% of ZINWA's historical gauge stations are functional and providing near-real-time data, thereby generating significant spatial deficiencies in observational coverage across numerous high-risk sub-catchment zones. The flood forecasting techniques presently utilized by ZINWA are largely based on deterministic conceptual rainfall-runoff models and empirical rating curve extrapolations, which necessitate comprehensive calibration datasets, exhibit suboptimal performance under non-stationary conditions, and provide limited probabilistic uncertainty quantification, functionalities that are crucial for risk-based early warning communication to emergency managers and the general populace [6, 7].

The global proliferation of machine learning (ML) and deep learning (DL) approaches in hydrology and natural hazard science has created a transformative opportunity to improve flood prediction capability in monitoring-limited environments such as Zimbabwe. Data-driven models can exploit heterogeneous data streams, including satellite rainfall products, remotely sensed land surface characteristics, and sparse in-situ records, without requiring the dense observational networks demanded by calibrated physics-based models [8, 9]. The increasing operational availability of high-quality, free satellite datasets, particularly NASA's Global Precipitation Measurement (GPM) mission and the ESA Copernicus Sentinel-1 Synthetic Aperture Radar programme, has further reduced the data barriers that historically constrained hydrological modelling in sub-Saharan Africa [10, 11]. Deep learning architectures, and Long Short-Term Memory (LSTM) networks in particular, have demonstrated the capacity to model complex nonlinear relationships between environmental forcing variables and hydrological response across diverse catchment settings without explicit process formulation [12, 13].

Despite these advances, the application of modern deep learning flood prediction frameworks to Zimbabwean and broader Southern African catchment systems remains limited. Most published studies have focused on data-rich basins in Asia, Europe, or North America, leaving significant gaps in the literature regarding model performance characteristics, optimal data requirements, and the translational potential of DL-based prediction for operational early warning in resource-constrained African institutional contexts [14, 15]. Furthermore, few studies have simultaneously addressed multiple flood typologies, cyclone-driven fluvial inundation, convective flash flooding, and urban drainage failure, within a unified deep learning architecture applicable across contrasting catchment types within a single national jurisdiction. The absence of model interpretability frameworks further limits practitioner uptake of ML-based flood tools in contexts where emergency managers require transparent, physically grounded justification for forecast-driven resource allocation decisions [16].

This study addresses these gaps by developing, training, and evaluating a comprehensive deep learning framework for flood prediction and early warning across the Save, Manyame, and Mazowe catchments of Zimbabwe. The specific contributions of this study are fourfold. First, a harmonized multi-source environmental feature dataset spanning 1985 to 2023 is compiled for three contrasting Zimbabwean catchments, combining satellite, geospatial, and in-situ data streams in a unified predictive architecture. Second, a novel CNN-LSTM hybrid model that jointly extracts spatial environmental patterns and temporal hydrological dynamics is proposed, trained, and compared against six benchmark algorithms under a consistent experimental protocol. Third, SHAP-based feature attribution is applied to provide physically interpretable insights into flood risk drivers that can support operational communication to ZINWA and Civil Protection Unit decision-makers. Fourth, a probabilistic alert threshold framework translating model outputs into a four-level early warning protocol is proposed for operational integration into Zimbabwe's existing flood monitoring institutional architecture.

While CNN-LSTM models have been applied in other regions, their use in Southern Africa remains rare. The novelty of this study lies in three areas: (1) the development of a harmonised multi-source dataset specifically tailored to Zimbabwe’s hydroclimatic conditions and data limitations; (2) the design of a hybrid CNN-LSTM architecture that explicitly fuses gridded spatial features (terrain, land cover, and satellite rainfall fields) with catchment-specific temporal dynamics; and (3) the translation of model outputs into a practical, institutionally aligned early warning protocol for ZINWA and the Civil Protection Unit. By focusing on three contrasting catchments within a single national context, this work provides the first comprehensive demonstration of how deep learning can bridge the gap between research and operational flood management in Zimbabwe.

Study Area

Three catchments administered by the Zimbabwe National Water Authority were selected based on their contrasting hydroclimatic settings, flood typologies, data availability, and collective coverage of Zimbabwe's principal flood-risk landscape. Table 1 presents their key physical and hydrological characteristics.

Table 1. Physical and hydrological characteristics of the three study catchments.

Catchment	Area (km²)	Mean Annual Rainfall (mm)	Flood Events (1985–2023)	Primary Flood Mechanism	ZINWA Sub-Region
Save	106,000	650–1,800	218	Cyclonic / Fluvial	Save Sub-Region
Manyame	16,540	750–1,050	174	Urban Flash	Manyame Sub-Region
Mazowe	38,400	700–1,100	196	Fluvial / Seasonal	Mazowe Sub-Region

Save Catchment

The Save Catchment is Zimbabwe's largest ZINWA-administered sub-basin, draining approximately 106,000 km² across the eastern and south-eastern portions of the country before discharging into the Indian Ocean in central Mozambique. It encompasses the Eastern Highlands, including the Chimanimani, Nyanga, and Vumba mountain ranges, where orographic enhancement drives annual rainfall totals regularly exceeding 1,800 mm on windward slopes, contrasting sharply with sub-200 mm totals in the arid lowveld zones of the south-west [4]. This extreme rainfall gradient, combined with the catchment's steeply dissected topography, shallow granite-derived soils, and large extent of communal lands subject to progressive vegetation degradation, makes the Save the most flood-prone river system in Zimbabwe. The catchment experienced Zimbabwe's two most destructive recent flood disasters, Tropical Cyclone Eline (February 2000) and Tropical Cyclone Idai (March 2019), both of which produced catastrophic inundation, landslides, and infrastructure destruction in the mountainous headwaters and broad alluvial flood plains of the middle and lower Save [3]. A dataset of 218 documented flood events spanning 1985 to 2023 was compiled for this study from ZINWA gauge records, EM-DAT global disaster database entries, Civil Protection Unit incident reports, and Sentinel-1 SAR flood footprint archives.

Manyame Catchment

The Manyame Catchment covers approximately 16,540 km² in north-central Zimbabwe, encompassing the capital city of Harare, its rapidly expanding peri-urban fringe, and an agricultural hinterland extending south-west toward the Sanyati Basin margin. The catchment is head watered by the Mukuvisi, Marimba, and Manyame rivers and is regulated by a cascade of impoundments, Lake Chivero, Lake Manyame, and Harava Dam, that supply Harare's population of approximately 2.7 million with domestic water [5]. Mean annual rainfall ranges from 750 to 1,050 mm, concentrated in the November to March wet season, with a high frequency of convective events capable of generating intensity-duration combinations that exceed the drainage capacity of Harare's aging storm-water infrastructure. Flash floods in the Manyame Catchment area are primarily urban and peri-urban events. These are caused by a combination of heavy rainfall, the growing number of hard surfaces due to Harare's rapid growth of informal settlements, and the ongoing blockage of culverts and drainage systems. The Budiriro, Glen View, Hopley, and Caledonia suburbs are subject to near-annual inundation of informal housing situated within the Mukuvisi and Marimba floodplains [38]. A dataset, including 174 flood events from 1985 to 2023, was created to train and validate the model.

Mazowe Catchment

The Mazowe Catchment drains approximately 38,400 km² across north-eastern Zimbabwe into the Zambezi River system, encompassing a diverse landscape gradient from the productive commercial farming zones of Mashonaland Central province through the communal lands of the mid-Mazowe valley to the hot, semi-arid Zambezi escarpment and valley floor. Major tributaries include the Nyadiri, Ruya, Nyagui, and Musengezi rivers. Mean annual rainfall ranges from 700 mm in the north-east escarpment zone to over 1,100 mm in the wetter south-western headwaters near Bindura and Shamva [6]. Flooding within the Mazowe Catchment is primarily characterized by fluvial and seasonal patterns, primarily instigated by extended multi-day rainfall occurrences linked to tropical disturbances, the Inter-Tropical Convergence Zone's influence, and the interplay of synoptic-scale troughs with orographic features in the eastern headwaters. The catchment area witnessed substantial flooding during the 1996/97, 2000/01, 2006/07, 2011/12, and 2021/22 seasons, resulting in documented damage to smallholder agriculture, rural road bridges, and school infrastructure. To support model development, a dataset of flood events was created, including 196 instances from 1985 to 2023.

Flood Prediction Using Machine Learning: A Global Overview

The systematic use of machine learning for flood prediction has seen significant growth since the mid-2010s. This increase is due to better access to computing power, the availability of open Earth observation data, and rapid advancements in algorithm development. Mosavi et al. [14] conducted a thorough review of machine learning studies on flood prediction. They found that Random Forest, Support Vector Machine, and Artificial Neural Network models were the most used. They also concluded that ensemble methods and deep learning approaches usually perform better than traditional statistical models when there is enough training data. The review highlighted that a lack of data in developing countries is the main barrier to using models in different places and for practical applications. This finding is particularly relevant to Zimbabwe, where monitoring is limited.

Random Forest (RF), as proposed by Breiman [17], generates predictive ensembles by combining the outputs of numerous bootstrapped decision trees. This approach provides resilience against noisy and correlated features, incorporates built-in feature importance estimation, and exhibits considerable resistance to overfitting. Consequently, these characteristics have rendered RF the most prevalent baseline algorithm in global flood susceptibility mapping research [18, 19]. Tyralis et al. [20] examined the use of random forests (RF) in water science, concluding that their inherent robustness to hyperparameter tuning makes them particularly advantageous for practical applications in resource-constrained settings. XGBoost, as presented by Chen and Guestrin [21], improves upon gradient boosting by incorporating a regularized objective function and parallel processing, and it has repeatedly demonstrated superior performance in predicting tabular data and assessing flood susceptibility [22, 23]. Choubin et al. [24] provided evidence of XGBoost's superiority over discriminant analysis, classification trees, and SVM in the context of flood susceptibility mapping, attributing this advantage to its ability to model higher-order feature interactions.

Support Vector Machines (SVMs) with radial basis function kernels have demonstrated effectiveness in flood susceptibility modelling, particularly with moderate-sized datasets and high-dimensional feature spaces [25, 26]. Nonetheless, SVMs are ill-equipped for the large datasets generated by multi-year, catchment-scale environmental monitoring. Furthermore, their inherent inability to provide probabilistic outputs constrains their applicability within threshold-based early warning systems, such as the one proposed in this research.

Deep Learning in Hydrological Forecasting

The growing use of Long Short-Term Memory (LSTM) networks as the main deep learning method for hydrological time series modelling has significantly changed data-driven hydrology in the last decade. LSTM networks, first introduced by Hochreiter and Schmid Huber [27], help to solve the vanishing gradient problem, which is a common issue in traditional recurrent neural networks. They achieve this through gated memory cell mechanisms, which selectively preserve, modify, or eliminate information across prolonged temporal sequences. Kratzert et al. [28] validated LSTM as a viable alternative to calibrated process-based rainfall-runoff models in a seminal study encompassing numerous catchments within the CAMELS benchmark dataset. Their research revealed that LSTM models could achieve performance levels comparable to or surpass those of expertly calibrated conceptual models, without necessitating explicit parameterization of hydrological processes.

Subsequent studies have confirmed the utility of LSTM for flood prediction at operational lead times. Le et al. [29] demonstrated that LSTM flood stage forecasts achieved substantially lower mean absolute error than ARIMA and shallow neural network benchmarks at lead times of up to 24 hours in a Vietnamese river basin. Hu et al. [30] established consistent LSTM performance advantages over multi-layer perceptrons and statistical regression models for rainfall-runoff simulation across multiple Chinese catchments, with advantages increasing at longer forecast lead times. Kratzert et al. [31] extended the LSTM rainfall-runoff modelling paradigm to demonstrate regional transferability training a single model across large and diverse catchment populations with implications for the data-limited contexts characteristic of sub-Saharan African river systems.

Convolutional Neural Networks, which are typically used for image classification, have been adapted to extract spatially structured predictive features from gridded environmental datasets. These datasets include precipitation fields, derivatives of Digital Elevation Models (DEMs), and satellite imagery [32]. Wang et al. [33] applied a convolutional approach to flood susceptibility mapping using satellite-derived inputs and obtained AUC-ROC values exceeding 0.97. Hybrid CNN-LSTM architectures, which combine CNNs for extracting spatial features with LSTMs for modelling time, have shown promise in flood prediction. This is especially true when the prediction involves spatially distributed factors and a catchment's response over time [34, 35]. These hybrid architectures form the basis for the model developed in this study.

Flood Prediction and Early Warning in Sub-Saharan Africa

The existing research on machine learning (ML)-based flood prediction in sub-Saharan Africa is limited, especially considering the region's significant flood risk. This is due to a lack of observational data, limited research resources, and institutional barriers to accessing data. Mugume et al. [36] reviewed the development of flood early warning systems across sub-Saharan Africa. They identified network problems, data transmission failures, and fragmented institutions as major obstacles to practical implementation. This is particularly relevant to the Zimbabwean context described in Section 1. Maposa et al. [37] studied extreme rainfall patterns in the Limpopo River Basin, which borders Zimbabwe. Their findings provided statistical evidence of changing patterns in rainfall that causes floods, which directly affects the time needed for ML training datasets. Chikodzi et al. [38] examined the geographical distribution of flood exposure in Zimbabwe using historical disaster records and satellite data. They found that the Save, Manyame, and Mazowe catchments together account for over 65% of Zimbabwe's reported flood-affected population over the past thirty years.

Mugabe et al. [39] examined hydrological research opportunities in the Save and Runde catchments, identifying the interaction of tropical cyclone rainfall with the catchment's granitic geology and shallow soils as the primary driver of rapid runoff generation and flash flood onset. Maidment et al. [40] demonstrated that GPM-family satellite precipitation products provide adequate accuracy for hydrological modelling in East and Southern Africa at basin scales, supporting the use of GPM IMERG as a primary precipitation input in data-scarce settings such as the ZINWA gauge network gaps that characterize portions of the study catchments.

Model Interpretability and Operational Integration

The extensive implementation of machine learning-driven flood prediction within operational contexts has been hampered by the perceived lack of transparency inherent in intricate model architectures. This opacity restricts the capacity of hydrological professionals and emergency management personnel to scrutinize, verify, and ultimately place their trust in the outputs generated by these models [16, 41]. SHAP (SHapley Additive exPlanations), a method introduced by Lundberg and Lee [42], offers a solution to this challenge. It furnishes feature attribution values derived from cooperative game theory, thereby adhering to established formal principles of local accuracy, consistency, and missingness. SHAP-based feature importance analysis has been applied in a growing number of flood and natural hazard ML studies, consistently producing feature rankings that align with domain-expert knowledge and that are more robust to feature correlation than permutation-based importance metrics [22, 43].

Goetz et al. Interpretability, as [44] posited, is an essential facet of natural hazard decision support models, rather than a supplementary feature, because emergency managers, who are directly responsible for public safety, are unlikely to act on model predictions that lack explanation. This perspective is especially relevant within Zimbabwe, where ZINWA and the Civil Protection Unit's technical personnel possess limited expertise in deep learning methodologies; these individuals must effectively convey the scientific underpinnings of flood alerts to district administrators, community leaders, and international humanitarian organizations. Consequently, the SHAP attribution and alert threshold frameworks developed in this research are specifically intended to reconcile the complexity of model architecture with the practical communication needs of operational contexts.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

METHODOLOGY

Multi-Source Data Acquisition and Pre-Processing

Table 2 summarizes the eight data categories integrated into the prediction framework. All datasets were reprojected to the WGS84 / UTM Zone 36S coordinate reference system and resampled to a harmonized 500-metre spatial resolution using bilinear interpolation for continuous fields and nearest-neighbour resampling for categorical variables. Temporal harmonization was achieved by aggregating all time-varying inputs to a common daily time step consistent with the ZINWA streamflow gauge reporting frequency.

Table 2. Multi-source environmental datasets integrated into the deep learning flood prediction framework.

Dataset	Variable(s)	Spatial Resolution	Temporal Resolution	Source
GPM IMERG V06	Precipitation	0.1°	30-minute	NASA GES DISC
SRTM DEM V3	Elevation, Slope, TWI, SPI	30 m	Static	USGS Earth Explorer
MODIS MOD13A1	NDVI	500 m	16-day composite	NASA LP DAAC
MODIS MCD12Q1	Land Use / Land Cover	500 m	Annual	NASA LP DAAC
Sentinel-1 GRD SAR	Inundation Extent (VV/VH)	10 m	Event-based	ESA Copernicus Hub
ZINWA Gauge Network	Streamflow, Water Level	Point (38 gauges)	Daily	ZINWA, Harare
ISRIC Soil Grids 250m	Soil Permeability, Clay Content	250 m	Static	ISRIC World Soil Info.
IBTrACS v04r00	Cyclone Track / Intensity	Point / Track	6-hourly	NOAA / WMO

GPM IMERG Final Run Version 06 data were obtained from the NASA Goddard Earth Sciences Data and Information Services Centre (GES DISC) for the period January 2000 to December 2023 at 0.1-degree spatial resolution and 30-minute temporal resolution. For the 1985 to 1999 portion of the study period, TRMM 3B42 Version 7 data were used as a structurally compatible predecessor, following the approach of Maidment et al. [40]. Bias correction against ZINWA and MSD gauge records used the quantile delta mapping method [45], with station-specific correction factors computed separately for each of the three catchments to account for orographic and regional rainfall gradient effects. SRTM Version 3 DEM data at 1-arcsecond (~30 m) resolution were downloaded from the USGS Earth Explorer portal and hydrologically conditioned using the Priority-Flood depression-filling algorithm prior to flow accumulation analysis [46]. Seven terrain derivatives were computed within SAGA GIS [47]: slope gradient (degrees), aspect (degrees), profile curvature, planar curvature, Topographic Wetness Index ($TWI = \ln(a/\tan\beta)$), Stream Power Index, and drainage density. Distance to the nearest delineated river channel, a key predictor of proximity-based inundation risk, was computed as Euclidean distance from each grid cell centroid to the ZINWA-derived hydrographic network, supplemented by channels delineated at a 1-km^2 flow accumulation threshold.

MODIS MOD13A1 Version 6 16-day NDVI composites at 500 m and annual land cover classifications from MCD12Q1 Type 1 were sourced from the NASA LP DAAC for 2001 to 2023. Pre-MODIS era cover for 1985 to 2000 was estimated from Landsat 4/5 TM classification products in the Google Earth Engine archive. Sentinel-

1 Ground Range Detected SAR imagery in VV and VH dual-polarization was sourced from the ESA Copernicus Open Access Hub for all documented flood events with concurrent cloud-free acquisition opportunities, yielding 127 validated inundation footprints for training label assignments. IBTrACS Version 04r00 tropical cyclone track and intensity data were downloaded from NOAA and used to construct a cyclone proximity index for each grid cell and time step as the product of the inverse squared distance to the nearest active tropical disturbance centre and the simultaneous maximum sustained wind speed, providing a continuous measure of cyclone forcing particularly relevant to the Save Catchment.

Feature Engineering and Label Construction

A total of 44 input features were derived from the eight raw data streams. Antecedent precipitation indices were calculated at five accumulation windows, 24 hours, 48 hours, 72 hours, 7 days, and 30 days, to capture the progressive depletion of catchment storage capacity prior to flood onset, consistent with the antecedent moisture control on variable source area runoff generation documented in Zimbabwe's granite-dominated catchments [48]. Streamflow anomalies at 7-day and 30-day moving average windows were computed from ZINWA daily gauge observations at 38 active stations distributed across the three catchments, expressed as standardized deviations from the long-term seasonal climatological mean discharge at each gauge to isolate above-normal flow conditions from the seasonal signal. Soil moisture proxy indices combining antecedent rainfall, TWI, and soil permeability were additionally derived to provide a spatially continuous representation of pre-event catchment wetness in areas distant from active gauge stations.

Flood occurrence labels were assigned at the grid cell level and daily time step using a three-source evidence combination approach: (i) ZINWA gauge-based flood event records; (ii) Sentinel-1 SAR inundation extent polygons co-registered to the study grid; and (iii) Civil Protection Unit incident database records spatially referenced to affected community locations. A grid cell was assigned a positive flood label if any of these three sources confirmed inundation within a ±3-day temporal window centred on the labelling date. Non-flood observations were drawn from periods at least 21 days removed from any documented flood event to minimize label contamination from antecedent moisture conditions associated with transitional pre- and post-flood states. The combined labelled dataset comprised 592,160 observations, of which 21.3% carry positive flood labels, reflecting the natural class imbalance of flood occurrence in a multi-year, multi-catchment dataset. Class imbalance was addressed during model training through inverse-frequency class weighting.

Deep Learning Architecture: CNN-LSTM Hybrid

The proposed CNN-LSTM hybrid architecture is designed to simultaneously exploit the spatial structure of gridded environmental inputs and the temporal dynamics of sequential hydrological data, two complementary dimensions of flood generating processes that no single-modality architecture can fully represent. The CNN component processes gridded spatial inputs (GPM rainfall fields, DEM terrain derivatives, NDVI, land cover) as multi-channel spatial arrays of dimensions H × W × C, where H and W are the spatial domain dimensions and C is the number of input channels. Two convolutional blocks, each comprising a 3×3 convolutional layer, batch normalization, and ReLU activation followed by 2×2 max-pooling, progressively extract hierarchical spatial features from 32 to 64 filter channels. The flattened CNN output vector is concatenated with a parallel temporal input vector comprising scalar time series features from ZINWA gauge observations, antecedent rainfall indices, and the cyclone proximity index, organized into 72-step (72-day) sliding windows that encompass sufficient temporal context for multi-week antecedent moisture dynamics and cyclone approach trajectories. The concatenated spatial-temporal representation is passed to a two-layer stacked LSTM with 256 and 128 units respectively, followed by dropout regularization and a dense output layer producing a sigmoid-normalized flood probability in [0, 1]. Table 3 provides the full layer-by-layer architecture specification.

Table 3. CNN-LSTM hybrid architecture specification.

Layer	Type / Operation	Configuration	Output Shape	Purpose
1	Conv2D + BN + ReLU	32 filters, 3×3 kernel	H×W×32	Spatial feature extraction

2	MaxPool2D	2×2 pool size	H/2×W/2×32	Dimensionality reduction
3	Conv2D + BN + ReLU	64 filters, 3×3 kernel	H/2×W/2×64	High-level spatial features
4	MaxPool2D + Flatten	2×2 pool size	1D spatial vector	Prepare for temporal input
5	Concatenation	CNN output + scalar time series	Combined feature vector	Multi-source fusion
6	LSTM Layer 1	256 units, return sequences	T×256	Long-range temporal modeling
7	LSTM Layer 2	128 units	128	Sequence summarization
8	Dropout	Rate = 0.30	128	Regularization
9	Dense + ReLU	64 neurons	64	Nonlinear transformation
10	Dense + Sigmoid	1 neuron	1 (probability)	Flood probability output

All datasets were meticulously harmonized to facilitate consistency and reproducibility. The spatial resolution of the datasets was harmonized to 500 m grid cells using bilinear interpolation for continuous datasets and nearest-neighbour resampling for categorical data. All datasets were temporally harmonized by consolidating them into daily resolution based on the daily resolution gauge reports by ZINWA. In the years 1985-1999, the TRMM 3B42 was used as an appropriate historical dataset equivalent to the GPM IMERG. Quantile Delta Mapping bias adjustment was applied to the TRMM 3B42 data separately for each basin to address regional variations in rainfall due to topographical effects. Terrain indices were generated from the SRTM DEM using the SAGA GIS, while the SAR images of the Sentinel-1 satellite were only selected on clear days associated with the observed floods.

Training used the Adam optimizer [49] with an initial learning rate of 10^{-3}, binary cross-entropy loss function, mini-batch size of 512, and early stopping with patience of 15 epochs monitored on the validation set loss. Learning rate reduction on plateau (factor 0.5, patience 7 epochs) was applied to stabilize convergence. All continuous features were standardized to zero mean and unit variance using statistics computed on the training set and applied consistently to the validation and test sets to prevent data leakage. The complete training pipeline was implemented in Python 3.11 using TensorFlow 2.14 and Keras, with NumPy, Pandas, scikit-learn, and the SHAP library for feature attribution. All experiments were conducted on a single NVIDIA RTX 3090 GPU.

Benchmark Models

Six benchmark algorithms were evaluated against the proposed CNN-LSTM: Logistic Regression, Decision Tree, Random Forest [17], XGBoost [21], Support Vector Machine with RBF kernel [50], and a standalone LSTM with identical architecture to the LSTM component of the hybrid but receiving only the temporal scalar input vector without CNN-extracted spatial features. Logistic Regression and Decision Tree serve as simple linear and nonlinear baselines. Random Forest and XGBoost are the most effective traditional ensemble methods for predicting floods using tabular data. Support Vector Machines (SVM) are the standard benchmark for kernel machines. The standalone LSTM isolates the contribution of spatial CNN feature extraction by comparison with the CNN-LSTM. All benchmark models were trained on the same 80/20 stratified train/test split and evaluated under identical experimental conditions to ensure fair comparison. Hyperparameter optimization for all models used five-fold cross-validated grid search with AUC-ROC as the selection criterion, as detailed for each algorithm in Supplementary Materials Table S1.

Evaluation Protocol and Spatial Cross-Validation

Model performance was assessed using five metrics: Overall Accuracy, Precision, Recall, F1-Score, and AUC-ROC. F1-Score and AUC-ROC were designated primary evaluation criteria given the class imbalance in the dataset and the asymmetric costs of false negative predictions in an early warning context, a missed flood warning carries substantially greater human cost than a false alarm. Brier Skill Score was additionally computed to evaluate probabilistic calibration of models producing continuous output probabilities. Statistical significance of pairwise AUC differences was tested using the DeLong test at a 95% confidence level.

Spatial cross-validation assessed out-of-sample transferability by rotating the held-out test catchment across all three permutations of the three-basin pool: training on Manyame + Mazowe, evaluating on Save; training on Save + Mazowe, evaluating on Manyame; and training on Save + Manyame, evaluating on Mazowe. This design tests whether models trained in two Zimbabwean catchments can generalize to a third with distinct flood typology and catchment characteristics, a direct proxy for the operational requirement to extend a pre-trained model to inadequately gauged sub-catchments within the ZINWA network.

SHAP Feature Attribution

SHAP values were computed for the CNN-LSTM and XGBoost models using the Deep Explainer and Tree Explainer implementations respectively in the SHAP Python library Version 0.44 [42]. For each prediction in the held-out test set, SHAP values quantify the marginal contribution of each input feature to the deviation of that prediction from the model's baseline flood probability. Global feature importance was determined by calculating the mean absolute SHAP value across all predictions within the test set, thereby establishing a rank order of feature influence that adheres to the SHAP consistency and efficiency principles. To discern nonlinear feature effects and synergistic interactions among the most influential predictors, SHAP dependence plots and pairwise interaction matrices were examined. This analysis specifically focused on the interaction between antecedent rainfall and TWI, which serves as a proxy for the variable source area runoff mechanism observed in Zimbabwe's granite-dominated catchments, as documented in [48].

Early Warning Threshold Framework

A four-tiered probabilistic alert threshold matrix was developed to convert the CNN-LSTM's continuous flood probability output into operationally useful early warning levels. These thresholds were calibrated using the empirical distribution of flood probabilities associated with confirmed flood observations within the test dataset. The levels were then defined to minimize the combined costs of false negatives (missed warnings) and false positives (unnecessary evacuations), considering the asymmetric consequence profiles characteristic of sub-Saharan African humanitarian response scenarios [36]. The four levels, Advisory, Watch, Warning, and Emergency, are aligned with the terminology and institutional responsibilities of Zimbabwe's Civil Protection Act and ZINWA's operational flood monitoring mandate.

RESULTS

Comparative Model Performance

Table 4 presents the classification performance metrics for the seven models assessed using the pooled held-out test set, which included 118,760 observations. These observations were selected from all three catchments, reflecting their respective dataset sizes. The CNN-LSTM hybrid model demonstrated superior performance across all five-evaluation metrics: an accuracy of 95.9%, a precision of 95.3%, a recall of 94.8%, an F1-score of 95.0%, and an AUC-ROC of 0.981. The standalone Long Short-Term Memory (LSTM) model, which achieved an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.971, demonstrated the utility of temporal sequence modelling in hydrological flood prediction, even in the absence of Convolutional Neural Network (CNN) spatial enhancement. XGBoost ranked third (AUC-ROC 0.964), substantially outperforming Random Forest (AUC-ROC 0.943) and SVM (AUC-ROC 0.924). Logistic Regression, with an AUC-ROC of 0.814, demonstrated the least predictive power. The results imply that a linear classifier proved insufficient for modelling the complex, nonlinear interactions inherent in the environmental factors influencing flooding, especially considering the varied catchment characteristics observed across Zimbabwe. Model performance was rigorously evaluated using both the pooled held-out test set and spatial cross-validation to ensure robust generalisation. All seven models were trained and tested under identical conditions on an 80/20 stratified split. In addition to standard classification metrics, we computed the Brier Skill Score to assess probabilistic calibration, an important consideration for early warning applications. Statistical significance of differences in

AUC-ROC was tested using the DeLong test. This comprehensive validation framework goes beyond simple in-sample testing and directly addresses the need for reliable transferability in data-scarce African contexts.

Table 4. Comparative performance of all models on the pooled held-out test dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Logistic Regression	78.2	76.4	74.9	75.6	0.814
Decision Tree	82.7	81.1	79.8	80.4	0.841
Random Forest	90.4	89.1	88.5	88.8	0.943
XGBoost	93.1	92.4	91.0	91.7	0.964
SVM (RBF Kernel)	87.6	86.3	85.0	85.6	0.924
LSTM	94.2	93.3	92.7	93.0	0.971
CNN-LSTM (Proposed)	**95.9**	**95.3**	**94.8**	**95.0**	**0.981**

Note: Highlighted row (CNN-LSTM Proposed) indicates the best-performing model across all metrics.

DeLong test comparisons confirmed that the CNN-LSTM's AUC advantage over all other models was statistically significant at $p < 0.001$. The difference between LSTM and XGBoost was statistically significant ($p = 0.003$). In contrast, the difference between Random Forest and SVM was not statistically significant ($p = 0.17$). The CNN-LSTM's Brier Skill Score of 0.743, compared to 0.694 for the standalone LSTM and 0.671 for XGBoost, further confirms its superior probabilistic calibration, a property of particular importance for threshold-based early warning applications where the absolute value of forecast probability influences alert level assignment.

Spatial Cross-Validation and Model Transferability

Table 5 shows the AUC-ROC values from all three spatial cross-validation runs for the CNN-LSTM and XGBoost models. The CNN-LSTM model had an average out-of-sample AUC of 0.943 across the three runs, which is a decrease of only 4.0 percentage points compared to its in-sample performance. The model had the highest transfer AUC, 0.953, when trained on the Manyame and Mazowe catchments and tested on the Save Catchment. The model had learned to employ the cyclone proximity index from Manyame and Mazowe training periods, which may explain the improved result in the Save-dominated test set. In contrast, the model trained on Save and Mazowe and tested on the Manyame Catchment showed the weakest transfer performance, with an AUC of 0.931. This is consistent with the different flash flood generation mechanisms in Harare's peri-urban drainage system, which are not well represented in the training data from the other two catchments, which are mostly influenced by rivers and cyclones.

Table 5. Spatial cross-validation AUC-ROC results by training-test catchment permutation.

Training Catchments	Test Catchment	CNN-LSTM AUC	XGBoost AUC
Manyame + Mazowe	Save	0.953	0.947
Save + Mazowe	Manyame	0.931	0.916
Save + Manyame	Mazowe	0.944	0.938
Mean (all permutations)	—	**0.943**	**0.934**

SHAP Feature Attribution

Table 6 presents the top ten features by mean absolute SHAP value from the CNN-LSTM model, evaluated across the full pooled test set. The direction column indicates whether the feature exerts a positive (higher values increase flood probability) or negative (higher values decrease flood probability) average effect on model predictions.

Table 6. Top 10 features by mean absolute SHAP value: CNN-LSTM model (pooled test set).

Rank	Feature	SHAP Score	Direction	Data Source
1	72-hr Antecedent Rainfall	0.191	Positive	GPM IMERG
2	Distance to River Channel	0.159	Negative	SRTM DEM / ZINWA Hydro.
3	Terrain Wetness Index (TWI)	0.138	Positive	SRTM DEM
4	7-day Streamflow Anomaly	0.121	Positive	ZINWA Gauge Network
5	Elevation	0.097	Negative	SRTM DEM
6	Soil Permeability Index	0.083	Negative	ISRIC Soil Grids
7	Cyclone Proximity Index	0.072	Positive	IBTrACS / RSMC La Réunion
8	Slope Gradient	0.058	Positive	SRTM DEM
9	NDVI (16-day Composite)	0.047	Negative	MODIS MOD13A1
10	SAR Backscatter Intensity (VV)	0.034	Positive	Sentinel-1 GRD

The 72-hour antecedent precipitation proved to be the most significant determinant, as indicated by its SHAP score of 0.191 and its positive correlation. This finding corroborates the hypothesis that the accumulation of rainfall over a period of several days is the primary driver of saturating Zimbabwe's shallow, clay-dominated granitic soils prior to the onset of flooding, a phenomenon documented in reference [48]. Distance to river channel ranked second (0.159, negative direction), reflecting the strong hydraulic control of proximity to the delineated channel network on inundation probability in the broad alluvial flood plains of the middle and lower Save and in the low-gradient reaches of the Mazowe Valley. TWI ranked third (0.138, positive direction), confirming the topographic convergence control on runoff routing toward valley floors that is characteristic of Zimbabwe's granite-gneiss basement geology.

The cyclone proximity index ranked seventh (0.072, positive direction), with SHAP dependence plots revealing a highly nonlinear response concentrated in the upper tail of the index distribution, consistent with the threshold behaviour of cyclone-driven rainfall intensification as tropical systems approach the Save Catchment headwaters. SHAP interaction analysis identified a statistically significant positive synergistic effect between antecedent 72-hour rainfall and TWI: in grid cells with high TWI values, flood probability responded more steeply to accumulating antecedent rainfall than in lower-TWI cells, consistent with the variable source area theory of runoff generation in which saturated contributing areas expand nonlinearly from convergent landscape positions as catchment wetness increases [48]. NDVI exerted a consistent negative effect on flood probability (rank 9, SHAP score 0.047), reflecting the attenuation of runoff generation by vegetated surfaces through interception and enhanced infiltration, particularly relevant in the miombo woodland areas of the Mazowe headwaters where NDVI values vary substantially between wet and dry seasons.

Proposed Early Warning Threshold Framework

Table 7 presents the four-level probabilistic alert threshold framework, which is based on the CNN-LSTM flood probability output. The alert levels are defined by flood probability ranges. These ranges were calibrated using the real-world distribution of model outputs, which were derived from confirmed flood and non-flood observations in the test set. The thresholds were chosen to optimize the balance between sensitivity (reducing missed warnings) and specificity (reducing false alarms) at each level. The analysis was done with a focus on the unequal consequences that arose from humanitarian efforts in Southern Africa.

Table 7. Proposed four-level probabilistic early warning threshold matrix for Zimbabwe catchment operations.

Alert Level	**Flood Probability Threshold**	**Colour Code**	**Recommended Action**	**Responsible Agency**
Advisory	0.40 – 0.59	Yellow	Public communication; monitoring intensified	ZINWA / MSD
Watch	0.60 – 0.74	Amber	Activate response pre-positioning; community alerts	Civil Protection Unit
Warning	0.75 – 0.89	Orange	Evacuation of high-risk zones; road closures	CPU / District Administrators
Emergency	**≥ 0.90**	**Red**	**Full emergency declaration; search and rescue**	**National Emergency / CPU**

The four-tiered alert system was purposely designed to match with Zimbabwe's established institutional framework, as outlined in the Civil Protection Act. The thresholds were determined by analysing the empirical distribution of model outputs linked to confirmed flood occurrences, thereby weighing the significant human toll of unheeded warnings against the operational expenses incurred by false alerts For example, the "Emergency" level (probability ≥ 0.90) is triggered only when the model shows very high confidence, corresponding to the 95th percentile of historical major flood events. This framework can be directly integrated into ZINWA's daily flood bulletin. This would replace the current deterministic approach with probabilistic, easily understandable alerts, which emergency managers could then trust and use effectively.

The Advisory level (probability 0.40-0.59) is designed to trigger enhanced monitoring and public communication without activating expensive emergency response pre-positioning, minimizing the opportunity cost of false alarms at the lower sensitivity threshold. This structure ensures the system is both practical and trustworthy for operational use by ZINWA and the Civil Protection Unit.

DISCUSSION

Model Performance and Methodological Implications

The CNN-LSTM hybrid's AUC-ROC of 0.981 represents state-of-the-art flood prediction performance for a Southern African catchment system and substantially exceeds results reported in the limited available literature for comparable developing-world settings. Choubin et al. [24] reported AUC values of 0.87-0.92 for ensemble tree-based flood susceptibility models in Iranian catchments. Tehrany et al. [25] obtained AUC values of 0.88-0.93 for SVM-based flood mapping. The performance improvement demonstrated here is attributable to three methodological factors operating jointly: the richer multi-source feature space that integrates spatial, temporal, and meteorological information not available to single-source models; the CNN-LSTM's capacity to simultaneously extract spatial patterns from gridded inputs and temporal dynamics from sequential gauge and rainfall data; and the large and diverse training dataset assembled across three catchments spanning 38 years of documented flood history.

The performance differences seen across algorithm types, where deep learning performed best, followed by gradient boosting, kernel methods, and linear models, aligns with the overall findings of Mosavi et al.'s meta-analysis. [14] and with the theoretical expectations of each model class's representational capacity. The substantial performance gap between the standalone LSTM and the CNN-LSTM (AUC 0.971 vs 0.981) specifically isolates the contribution of CNN spatial feature extraction as an additive improvement beyond temporal modelling alone, confirming that the spatial distribution of rainfall, topography, and land cover within the catchment, not just their temporal evolution, carries predictive information not captured by point-scale time series modelling. This finding has practical implications for the design of operational DL flood prediction systems: the additional computational cost of CNN spatial feature processing is justified by a statistically significant performance improvement.

Transferability and Implications for Zimbabwe's National System

The spatial cross-validation results, with CNN-LSTM out-of-sample AUC consistently above 0.93 across all three catchment permutations, are of direct practical significance for Zimbabwe's national flood management architecture. ZINWA currently lacks the technical capacity to maintain and recalibrate independent catchment-specific models for each of its seven sub-regions. The demonstrated transferability of the CNN-LSTM across Zimbabwe's three largest and most hydrologically distinct flood-prone catchments suggests that a single national model, retrained annually with incoming GPM and ZINWA gauge data, could serve the operational forecasting needs of multiple sub-regions simultaneously, a finding aligned with the universal hydrological model concepts explored by Kratzert et al. [31] in the CAMELS context.

The relatively weaker transferability from fluvial-dominated Save and Mazowe training data to the urban flash flood environment of the Manyame Catchment (AUC 0.931 vs 0.953) highlights the continued importance of catchment-type representation in training data composition. Urban flash floods in peri-urban Harare are generated by sub-hourly convective bursts interacting with impervious surface drainage networks at spatial scales below the 500-metre harmonized input resolution used in this study. Improving model performance for urban catchments would likely require integration of Harare city drainage network topology, sub-500-metre spatial inputs from commercial SAR providers, and sub-daily rainfall inputs at finer temporal resolution than ZINWA's daily gauge reporting frequency, all areas warranting targeted investment in future system development.

Feature Importance and Hydrological Interpretation

The SHAP feature attribution results provide a physically coherent account of flood risk drivers in Zimbabwe's study catchments that aligns with established hydrological understanding and supports the credibility of the model for operational deployment. The dominance of antecedent 72-hour rainfall as the primary predictor is consistent with the pivotal role of multi-day precipitation accumulation in depleting catchment storage capacity in Zimbabwe's seasonally dry soils, which begin each wet season in a relatively desiccated state and require sustained rainfall to approach saturation thresholds above which surface runoff generation accelerates rapidly [48]. The strong positive interaction between antecedent rainfall and TWI identified in SHAP interaction analysis confirms the variable source area runoff mechanism in which topographically convergent areas with high TWI respond disproportionately to cumulative rainfall, consistent with Beven's [51] conceptual model of partial-area contributing to streamflow generation in headwater basins.

The negative effect of NDVI on flood probability is physically interpretable as the protective role of vegetated land surfaces in reducing net rainfall input to the catchment through canopy interception and transpiration, and in increasing soil infiltration capacity through root microporosity. This finding has direct management relevance in the Mazowe Catchment, where ongoing deforestation of miombo woodland in communal land areas has been documented as a contributor to increasing flood frequency and magnitude over the past three decades [39]. The relatively modest contribution of SAR backscatter intensity (rank 10) confirms its role as a concurrent flood detection signal rather than a precursory predictor, a finding consistent with the temporal mechanics of SAR-based inundation detection and with its appropriate use as a validation label source rather than a predictive feature in early warning applications.

Limitations and Future Research Directions

Several limitations of this study require clear acknowledgment. The binary flood/non-flood classification framing does not capture continuous inundation depth, areal extent, or duration information of direct relevance to damage estimation, insurance assessment, and graduated emergency response prioritization. The 500-metre harmonized spatial resolution may inadequately represent the micro-topographic drainage controls governing flood generation in peri-urban Harare, where critical infrastructure features such as blocked culverts and informal settlement drainage channels operate at sub-100-metre scales. The dependence on GPM IMERG as the primary precipitation input introduces quantifiable uncertainty in the Eastern Highlands, where satellite-gauge discrepancies can reach 30–50% of event total rainfall in complex orographic terrain [40], and further work on orographic bias correction specific to the Save Catchment headwaters is warranted.

The proposed early warning threshold matrix requires operational validation through structured simulation exercises with ZINWA and Civil Protection Unit personnel before deployment, and the alert-level cost assumptions embedded in threshold calibration should be reviewed with flood emergency response practitioners to ensure alignment with Zimbabwe-specific institutional and resource constraints. Future research should prioritize: (i) extending the framework to regression prediction of inundation depth and areal extent; (ii) integrating medium-range ECMWF ensemble NWP forecasts as time-varying meteorological inputs to extend warning lead times beyond the 24-72-hour horizon achievable with GPM near-real-time products; (iii) developing a computationally lightweight XGBoost surrogate model deployable on ZINWA operational servers without GPU infrastructure; and (iv) conducting co-design workshops with ZINWA hydrologists, MSD forecasters, and Civil Protection Unit coordinators to ensure that model output visualization and communication formats are operationally appropriate.

CONCLUSION

This study developed, trained, and evaluated a deep learning framework for flood prediction and early warning across the Save, Manyame, and Mazowe catchments of Zimbabwe- three contrasting flood-prone systems representing the country's principal hydroclimatic regimes and flood typologies. A multi-source feature dataset integrating GPM satellite rainfall, SRTM terrain derivatives, MODIS land surface products, Sentinel-1 SAR inundation observations, ZINWA gauge records, Soil Grids soil data, and IBTrACS cyclone track information was compiled and harmonized across a 38-year training period. Seven algorithms were evaluated, with the proposed CNN-LSTM hybrid architecture achieving the highest performance: accuracy 95.9%, F1-score 95.0%, AUC-ROC 0.981. Spatial cross-validation demonstrated strong transferability across catchments, with mean out-of-sample AUC of 0.943, confirming the viability of a unified national DL model as an alternative to separate catchment-specific systems. SHAP attribution identified antecedent 72-hour rainfall, distance to river channel, and Terrain Wetness Index as the dominant flood predictors, providing physically interpretable and operationally communicable insights for ZINWA and Civil Protection Unit practitioners. A four-level probabilistic alert threshold matrix, Advisory, Watch, Warning, and Emergency, was proposed for integration into Zimbabwe's operational flood monitoring framework.

The findings make three contributions to the flood prediction literature. First, they demonstrate that deep learning multi-source integration substantially outperforms conventional ensemble tree and kernel machine approaches for flood prediction in Southern African catchment systems, with performance advantages that are statistically significant and operationally meaningful. Second, they show that strong model transferability is achievable across Zimbabwe's diverse catchment types, supporting the practical and institutional case for a national rather than site-specific DL prediction infrastructure. Third, they provide a methodological template, combining multi-source data fusion, deep learning spatiotemporal modelling, SHAP interpretability, and probabilistic alert calibration, that is replicable for flood early warning development across other data-limited Southern African river basin jurisdictions.

The successful operational deployment of the proposed framework will require investment in ZINWA gauge network rehabilitation, particularly in the Mazowe and upper Save sub-catchments where monitoring gaps most constrain model input quality. Partnership between ZINWA, MSD, HIT, and international technical partners, including UNOCHA, the World Meteorological Organization's HydroSOS initiative, and the Southern African

Development Community's Climate Services Centre, is recommended to support the transition from research prototype to operational early warning tool.

REFERENCES

1. Nhamo, G., & Chikodzi, D. (Eds.). (2021). Cyclones in Southern Africa: Volume 1, Resilience, sustainability and future risk. Springer Nature. https://doi.org/10.1007/978-3-030-72235-6
2. CRED / UNDRR. (2022). Human cost of disasters: An overview of the last 20 years 2000–2019. Centre for Research on the Epidemiology of Disasters, Brussels.
3. Msowoya, K., Mathiszik, M., Phiri, C., & Mofor, N. (2020). Flooding in the Chimanimani district of Zimbabwe following Tropical Cyclone Idai. Natural Hazards and Earth System Sciences Discussions. https://doi.org/10.5194/nhess-2020-100
4. Mugabe, F. T., Hodnett, M. G., & Senzanje, A. (2007). Opportunities for hydrological research in the Runde and Save Catchments, Zimbabwe. Physics and Chemistry of the Earth, 32(15–18), 1107–1114. https://doi.org/10.1016/j.pce.2007.07.004
5. ZINWA. (2020). Annual report on water resources monitoring in Zimbabwe 2020. Zimbabwe National Water Authority, Harare.
6. Beven, K. J. (2012). Rainfall-runoff modelling: The primer (2nd ed.). Wiley-Blackwell.
7. Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. Journal of Hydrology, 375(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005
8. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
9. Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? Water Resources Research, 57(3), e2020WR028091. https://doi.org/10.1029/2020WR028091
10. Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The Global Precipitation Measurement Mission. Bulletin of the American Meteorological Society, 95(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1
11. Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: A fully automated processing chain. International Journal of Remote Sensing, 37(13), 2990–3004. https://doi.org/10.1080/01431161.2016.1192304
12. Hochreiter, S., & Schmid Huber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
13. Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54(11), 8558–8593. https://doi.org/10.1029/2018WR022643
14. Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536. https://doi.org/10.3390/w10111536
15. Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., & Bui, D. T. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modelling at Haraz watershed, northern Iran. Science of the Total Environment, 627, 744–755. https://doi.org/10.1016/j.scitotenv.2018.01.266
16. Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modelling. Computers & Geosciences, 81, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007
17. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
18. Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008
19. Rahmati, O., Pourghasemi, H. R., & Zeinivand, H. (2016). Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto International, 31(1), 42–70. https://doi.org/10.1080/10106049.2015.1041559

20. Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water, 11(5), 910. https://doi.org/10.3390/w11050910
21. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
22. Pham, B. T., Avand, M., Janizadeh, S., Phong, T. V., Al-Ansari, N., Ho, L. S., Das, S., Le, H. V., Amini, A., & Nguyen, H. Q. (2021). GIS-based ensemble soft computing models for flood susceptibility assessment. Applied Sciences, 11(4), 1805. https://doi.org/10.3390/app11041805
23. Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of the Total Environment, 651, 2087–2096. https://doi.org/10.1016/j.scitotenv.2018.10.064
24. Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of the Total Environment, 651, 2087–2096. https://doi.org/10.1016/j.scitotenv.2018.10.064
25. Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2015). Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stochastic Environmental Research and Risk Assessment, 29(4), 1149–1165. https://doi.org/10.1007/s00477-015-1021-9
26. Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. Water Science and Technology, 82(12), 2635–2670. https://doi.org/10.2166/wst.2020.369
27. Hochreiter, S., & Schmid Huber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
28. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) based deep learning. Hydrology and Earth System Sciences, 22(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018
29. Le, X.-H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387. https://doi.org/10.3390/w11071387
30. Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water, 10(11), 1543. https://doi.org/10.3390/w10111543
31. Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviours via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019
32. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
33. Wang, Y., Fang, Z., Hong, H., & Peng, L. (2020). Flood susceptibility mapping using convolutional deep learning approach with remote sensing evidential factors. Journal of Hydrology, 590, 125365. https://doi.org/10.1016/j.jhydrol.2020.125365
34. Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? Water Resources Research, 57(3), e2020WR028091. https://doi.org/10.1029/2020WR028091
35. Sat, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. Water Science and Technology, 82(12), 2635–2670. https://doi.org/10.2166/wst.2020.369
36. Mugume, I., Basalirwa, C., Waiswa, D., Reuder, J., Mesquita, M. D. S., Tao, S., & Ngailo, T. J. (2016). Comparison of parametric and nonparametric methods for analysing the bias of a numerical model. Atmosphere, 7(11), 135. https://doi.org/10.3390/atmos7110135
37. Maposa, D., Cochran, J. J., & Lesaoana, M. (2016). Investigating the occurrence and return periods of extreme floods in the Limpopo Province, South Africa. Natural Hazards and Earth System Sciences, 16(1), 137–148. https://doi.org/10.5194/nhess-16-137-2016

38. Chikodzi, D., Mutowo, G., & Murwendo, T. (2013). Assessing flood risk in Zimbabwe using GIS and remote sensing. Journal of Flood Risk Management, 6(2), 145–154. https://doi.org/10.1111/j.1753-318X.2012.01164.x
39. Mugabe, F. T., Hodnett, M. G., & Senzanje, A. (2007). Opportunities for hydrological research in the Runde and Save Catchments, Zimbabwe. Physics and Chemistry of the Earth, 32(15–18), 1107–1114. https://doi.org/10.1016/j.pce.2007.07.004
40. Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., Allan, R. P., Stein, T., Nkonde, E., & Senkunda, S. (2017). A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. Scientific Data, 4, 170063. https://doi.org/10.1038/sdata.2017.63
41. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
42. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). NeurIPS.
43. Li, Z., Liu, H., Luo, C., & Fu, G. (2021). Assessing surface water flood risks in urban areas using machine learning. Water, 13(24), 3520. https://doi.org/10.3390/w13243520
44. Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modelling. Computers & Geosciences, 81, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007
45. Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? Journal of Climate, 28(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1
46. Wang, L., & Liu, H. (2006). An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science, 20(2), 193–213. https://doi.org/10.1080/13658810500433453
47. Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development, 8(7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015
48. Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin, 24(1), 43–69. https://doi.org/10.1080/02626667909491834
49. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). https://arxiv.org/abs/1412.6980
50. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
51. Beven, K. J. (2012). Rainfall-runoff modelling: The primer (2nd ed.). Wiley-Blackwell.
52. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
53. Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS, Global ensemble streamflow forecasting and flood early warning. Hydrology and Earth System Sciences, 17(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013