

Ms. Anjali P. Kadam<sup>1</sup>Prof. Dr. Prakash  
Devale<sup>2</sup>

## Systematic Literature Review on Rain/Flood Prediction Using Machine Learning and Deep Learning Models



**Abstract:** - A literature review surveys what has already been written about a topic. It helps researchers understand the current state of knowledge, key debates, theories, and findings in the field. By reviewing existing studies, researcher can build a solid base for the own work. It shows that the research is grounded in established knowledge. It places the research in context, helping readers understand why the particular topic is important and how it fits into the broader field. Also, literature reviews help define and explain key terms and theories relevant to the study, providing a clear framework for analysis. It is a critical analysis that sets the stage for the research, shows the understanding of the topic, and guides future investigation. Hence, this paper focuses on the literature review for the proposed research from year 2019 to 2022.

**Keywords:** Machine learning, deep learning, Geological AI, Flood prediction, LSTM

### 1. Introduction

Floods as well as heavy rain fall incidents are key natural catastrophes leading to critical financial impairment and human struggles throughout the world. Appropriate auguration of rain fall and surges is very important for rapid alert systems and so disaster control. By way of the progression of computational power and so data availableness, Machine Learning (ML) and Deep Learning (DL) models have surfaced as potent options for geographical auguration. This methodical review is exploring different ML and DL models employed in rain as well as flood forecasting, featuring their strategies, functionality, and challenges.

### 2. Specific Literature

Author evaluated the recent GeoAI and machine learning functions in hydrological challenges. GeoAI has revealed positive aspects in nonlinear modeling, computational performance, incorporation of diverse data resources, superior legitimate prediction potential, and so the unraveling of innovative hydrological structures and procedures. A crucial disadvantage in several GeoAI models is the enough model positioning and so low physical interpretability, explainability, as well as model generalization [1].

In this study, statistical as well as , artificial intelligence (AI) strategies, i.e. multivariate linear regression (MLR), artificial neural network (ANN), support vector regression (SVR), wavelet SVR (WSVR), black widow optimization-SVR (BWO-SVR), and the manner of innovative gunner-SVR (AIG-SVR), were applied to reproduce the runoff procedure of the Karkheh catchment on a day-by-day time scale at the time of the statistical stage 2010-2020 [2].

The standard physical-based model is incapable to produce a large precision of prediction because of the inlayed disturbances, nonlinear as well as stochastic nature of hydrological data. Author reviewed the latest advancement of the AI strategies for river stream movement predicting through the last two decades. The initial facilities of forecasting model are produced from the neural network theory; supervised machine learning and then fuzzy logic principles [3].Reliable and well-timed auguration of surges can considerably reduce the loss of life as well as asset. Lately, numerous machine learning models are utilized for flooding prediction, featuring that their particular functionality is more suitable to common statistical versions. On the other hand, the existing models overlook the spatial functionality of surges, which drive flooding generation and so attentiveness [4].

<sup>1</sup>Research Scholar,  
Department of Computer Engineering,  
BVDU, COE, Pune, India

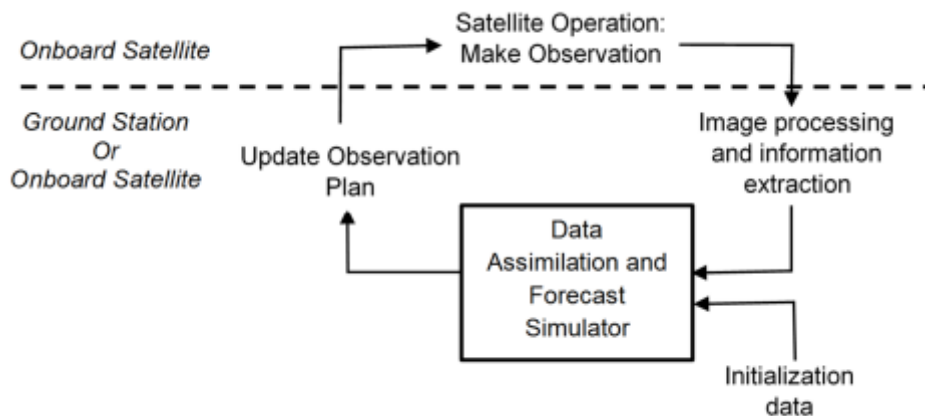
<sup>2</sup>Research Supervisor,  
Department of Information Technology,  
Bharati Vidyapeeth (Deemed to be) University,  
College of Engineering, Pune, Maharashtra, India

This assignment is designed to improve a flooding prediction system applying a collaboration of Machine Learning classifiers along with GIS approaches to be applied as a competent system for metropolitan administration and resilience scheduling. This strategy can create practical elements and risk indices for the incident of surges at the community tier that may be critical for setting out a long lasting technique for Smart Cities. The best performant Machine Learning model was a Random Forest, with a Matthew's Correlation Coefficient of 0.77 and so an Accuracy of 0.96 [5].

Because of the deceptive nature of flooding incidences, planning flooding auguration systems for early flooding alerts is challenging. This is having to the reality that early alerting system design requires a comprehensive understanding of a number of solutions.

Author recognized the Accuracy, Recall, and Receiver Operating Characteristics (ROC) results of 3 machine learning methods, namely Decision Tree, Logistic Regression, as well as , Support Vector Classification (SVR), were examined and investigated. Logistic Regression, in the event that likened with the various other two methods, provides more appropriate outcomes and gives superior effectiveness accuracy and recall. Furthermore, the Decision Tree performs better the Support Vector Classifier. Decision Tree worked moderately well because of its above-average accuracy as well as below-average recall scores. It is learned that Support Vector Classification worked inadequately with a modest proportions of dataset, with a recall score of 0, under average accuracy scores and a remarkably average roc score [6].

A data-driven model that uses simply every month rain fall data to achieve an exact result is intended to be economical options for countries around the world with modest information where technical discoveries haven't yet penetrated. The persistent break down triggered by flooding occurrences in the earlier research over various years is some other driving component behind this study. Author recommended two strategies to assimilate satellite viewed precipitation into hydrologic models in real-time to upgrade flooding forecasts, skipping two earlier limitations in this concept: sporadic satellite overpasses and so extended model run times. Multitude of modest geostationary satellites beats the initial hurdle by offering recurrent routes through an area of interest; on the other hand, these types of multitude necessitate co-ordination and scheduling to record anticipation data exactly where it is generally required to notify flooding forecasts. [7]. Refer Fig. 1 below. Provided this sparse prospect as well as , cost of assets to produce observations (data limits), it is often more effective to capture data which is expected to be practical instead of to first record an observation and afterwards identify its utility.



**Fig. 1: Adaptive sensing model [7]**

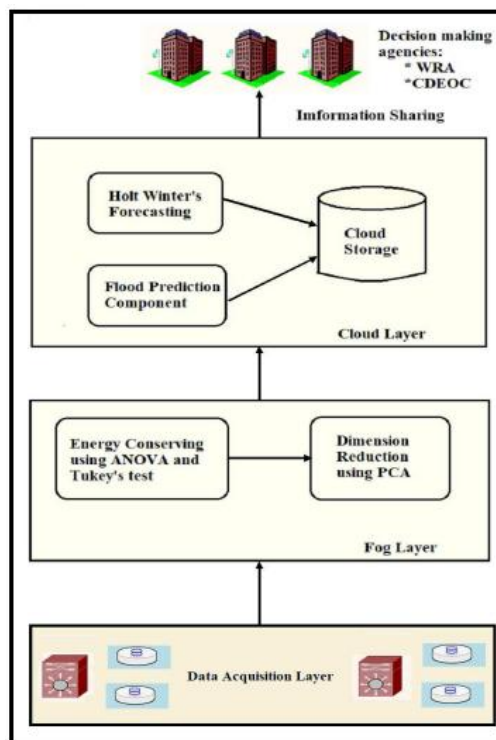
Nonetheless statistical models can effectively replicate and so forecast many of these incidents, it requires a long time to execute the linked computations, specifically two-dimensional considerable computations, which are not able to satisfy the requirements of crisis administration. Consequently, author suggests a technique of combining neural networks as well as statistical models that may replicate as well as determine aspects at increased concerns from metropolitan inundations and promptly forecast the depth of water deposition in these kinds of areas. Metropolitan flooding predicting analysis primarily aims at two versions of estimating depending on metropolitan flooding simulation models and so time series studies. Seeing that the 1970s, metropolitan flood models have been generated and designed swiftly and analysis on flood alert and auguration founded on

metropolitan flooding simulation models also initiated, and the concept is comparatively mature. However, a series of challenges, just like modest knowledge of hydrological and hydrodynamic procedures, intricate data requirements, as well as modeling issues, have limited the application of the models [8].

The outcomes of this analysis can immediately anticipate the sunken water range of the related concerns stage in the city depending on rainstorm data, which has the positive aspects of an elevated precision, a speedier response rate, and an even more correct auguration. Irrespective of its gains through the existing strategies, the suggested predictive modeling strategy for evaluation of flood susceptibility and problem is not rid of difficulties and constraints. A prevalent obstacle to all risk analysis methods for social influences is that, once outcomes are generated, it is complicated to reveal their systemic tendency and true concerns without access to a statistically adequately large percentage of catastrophes.

In the event that employing the risk analysis technique offered in this research, expert’s needs to also be mindful of its probable auguration prejudice and natural concerns. The probable bias of risk auguration extracted from the provided method may be germinated from 3 resources. Initial, the flooding threat map employed for this research may not be appropriate [9]. Beside flooding hazard as well as destruction data, a loss of adequate auguration of upcoming susceptibility elements could also results in propensity in model forecasts of flooding risk. In the proclaimed study, author utilized 2018 census data to obtain sociable susceptibility indications as proxies of sociable elements for flooding risk evaluation for upcoming years.

Author recommended an IoT-based energy efficient flooding auguration as well as predicting system. IoT sensor nodes are restricted when it comes to battery as well as memory, so the fog layer incorporates an energy-saving strategy depending on data heterogeneity to maintain the system’s power usage. Cloud storage is applied for effective storage. (Refer Fig.2 below)

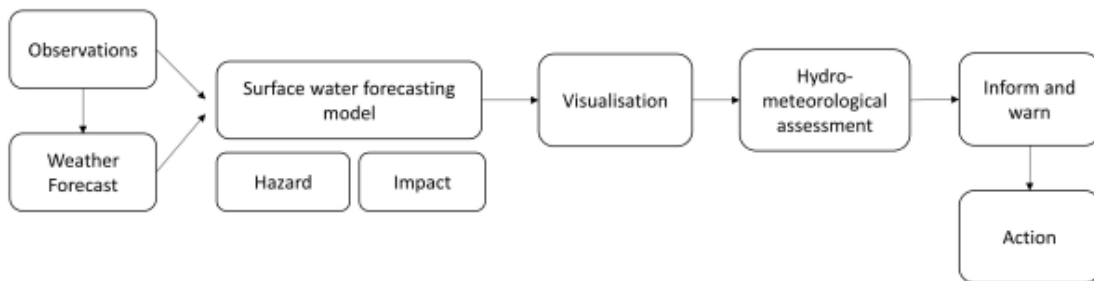


**Fig. 2: IoT based Cloud Framework [10]**

The geographical circumstances including temperature, humidity, rainfall, and water body parameters, i.e. water flow as well as water levels, are getting researched for India’s Kerala area to calibrate the flooding stages. PCA (Principal Component Analysis) strategy is applied at the fog layer for feature dimensionality decrease. To predict the flooding, the ANN (Artificial Neural Network) criteria are applied, and so the simulation approach of Holt Winter is applied to predict the potential flooding [10].

Author designed a hybrid deep learning (ConvLSTM) procedure including the predictive advantages of Convolutional Neural Network (CNN) as well as Long Short-Term Memory (LSTM) Network to model and examines a flood predicting model to estimate the upcoming incident of flooding situations. Extracted from precipitation data, the work explores a Flooding Index, in mode of a numerical manifestation, to record the progressive depletion of water resources eventually, applied in a flooding monitoring strategy to identify the length, intensity, and severity of any kind of overflow scenario. The outcomes shows functional use of ConvLSTM in correctly estimating flood index and its probable use in devastation administration and trouble minimization in the recent point of intense climate incidents [11]. It is suggested that in future study further user-friendly methods for flooding predicting be designed and other deep learning as well as machine learning methods be analyzed for flood predicting. The outcomes from this study can be establish as an assessment standard for the recently build models.

Author determined 3 methods to surface area water predicting implementing cutting edge meteorological predictions: empirical-based situations, hydrological estimations associated to pre-simulated effect cases, and then real-time hydrodynamic simulation. Examining functional cases of every strategy gives an occasion to uncover from international best practice to improve tailored, impact-based, surface area water estimations to assist recommended decision-making. Typical end-to-end structure is shown in Fig.3 below.



**Fig. 3: End-to-end surface water flood forecasting system [11]**

Even though the breakthrough of new meteorological as well as hydrological predicting features is encouraging, there is still a technological limit to the predictability of convective rainwater. To conquer this difficulty, author recommended that a change of the proven function of a water surge prediction is required, next to the advancement of interdisciplinary alternatives for interacting uncertainty and producing the ideal use of all obtainable data to boost readiness. [12].

Author proclaimed a model for short-term flooding prediction that incorporates IoT as well as ANN, where the prediction calculation is accomplished on a low power edge system. The model tracks timely rain fall and water level sensor data and incorporates the eventual correlative data for ahead-of-time auguration of flooding water levels employing long-short-term memory. The system can be integrated on battery-powered IoT systems. The outcome of assessing a model of the structure point out that the model is acceptable for timely flooding prediction. Author applied TensorFlow Lite for real-time ANN test of the sensor data on the edge device. TensorFlow Lite facilitates on-device ML inference with low latency and then a small binary size and is consequently appropriate for embedded and IoT systems.

Author examined the model for accuracy with ReLU, sigmoid as well as tanh initialization functions. The unit functions better in the event that applying ReLU initial function when compared with applying sigmoid or perhaps tanh constantly for estimations at distinct time frame. This is anticipated as ReLU converges better and, contrary to sigmoid, doesn't be affected from the disappearing slope issue [13]. Author offered an introduction of recent study activities in ensemble flooding predicting and examines awareness gaps as well as upcoming research possibilities, determined by an assessment of 70 papers concentrating on different facets of ensemble flooding predicting all across the domain. Potential research guidelines consist of chances to increase technological factors of ensemble flooding estimating, just like data compression approaches and strategies to concern for more resources of error, and expanding ensemble estimations for more parameters, for example flooding inundation, by employing methods including machine learning.

Additionally to this, author deduce that presently there is a call for progress in technological elements of flooding estimating, and also to link the gap among systematic research as well as , hydro-meteorological model advancement, and real-life flooding supervision utilizing probabilistic ensemble estimations, specifically with the aid of powerful interaction [14]. Difficulties lie ahead to generate ensemble flooding forecasting a prevalent tool employed for flooding administration or perhaps water management in overall. Initially, calculate a water surge parameters require to be widened to incorporate a water surge inundation range and level which are immediately affiliated to a water surge impact and often utilized in the water segment for a water surge risk/hazard analysis, as mentioned above. Further, forecast lead time requires to be widened to accommodate for many aspects of water management assignments.

Author presented an overview as well as information of latest improvements appropriate to ideas on atmospheric circumstances that come before excessive rain fall incidents, to the advancement of monitoring units of trusted hydro meteorological variables, and to the functional usage of climate and hydrological models toward the auguration of flash floods. For metropolitan conditions, in which the demand for high resolution simulations needs computationally steep units, query-based techniques tend to be discovered for the well-timed recuperation of pre-simulated flooding inundation predictions. Through the principle of the IoT, the considerable application of low-cost receptors opens possibilities from the point of view of denser monitoring potential. On the other hand, distinct environmental circumstances and unequal syndication of data as well as assets constantly directs to the usage of site-specific alternatives for flash flood projecting in the circumstance of rapid alert systems. [15].

Uttermost severe flash inundations may be identified as a primary purpose of large casualties and structure damage in various parts of the world like Pakistan, Malaysia, Philippines, Southern France, India, Bangladesh, China, Nepal, Canada, United States of America and others. Run offs can ruin enormous properties and individualized things within fraction of seconds. The constant deployment of early alert systems lead in predicaments in which diverse predicting methods, occasionally likewise structured on several criteria, were getting accomplished in parallel when it comes to the auguration of the comparable acknowledged flash flood occurrences. Developing several Artificial Intelligence Founded proficient methods to estimate the flash floods powerfully can be thought to be as non-engineering founded techniques. Authors have attempted the best to review and represent all the effective approaches that can be employed for the fast auguration of flash inundations. [16].

### 3. Method

For the systematic literature survey for the proposed study, we preferred the search Databases as Google Scholar, IEEE Xplore, ScienceDirect, Springer Link, Scopus. Further, for literature identification keywords used as “rainfall prediction machine learning,” “flood forecasting deep learning,” “hydrological modeling AI,” “flood early warning ML” During the search for appropriate literature, the inclusion criteria considered as Peer-reviewed articles (2019–2022), focus on ML/DL for rainfall/flood prediction, English language. We excluded Non-ML/DL based approaches, non-hydrological studies to scrutinize the best analysis papers.

Along with the literature search, we formulated research questions focusing on existing system studies.

#### 3.1. Research questions

To address the methodology of proposed research, few pilot research questions are formulated as below which can be evaluated at the stage of research methodology development:

RQ1. Which ML/DL methods are available in hydrological analysis?

RQ2. What are merits of existing system architectures?

RQ3. Which ML/DL algorithms available for studies?

RQ4. How can we focus over flood prediction DL methodologies?

**3.2. Database Search Process**

We followed systematic method to generate literature database. Initially, we identified significant electronic databases as: IEEE, Springer and Elsevier for publications search with phrase “hydrological modeling AI”. Following search string is developed and applied: (((Machine Learning\*) OR (Deep Learning) OR (Flood Prediction) OR))) AND ((rainfall prediction) OR (AI rainfall prediction\*)) The outcomes of the above query over the selected databases is given in Table 1, 2 and 3.

**Table 1: IEEE literature**

Input String	No. of papers			
	Conference	Journals	eBooks	Miscellaneous
Hydrological modeling AI	62	74	0	0
Machine Learning	89	106	1	0
Deep Learning	57	79	0	0
Flood Prediction	97	54	0	0
Rainfall prediction	32	21	0	0
AI rainfall prediction	16	7	0	0

**Table 2: Springer literature**

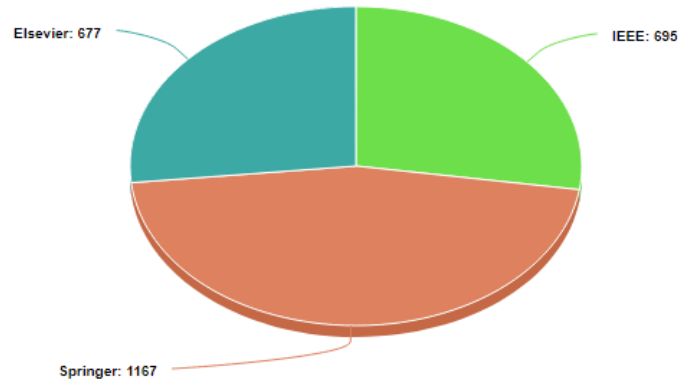
Input String	No. of papers			
	Conference	Journals	eBooks	Miscellaneous
Hydrological modeling AI	86	145	0	0
Machine Learning	124	231	0	3
Deep Learning	104	203	0	0
Flood Prediction	52	48	0	0
Rainfall prediction	74	68	0	0
AI rainfall prediction	9	20	0	0

**Table 3: Elsevier literature**

Input String	No. of papers			
	Conference	Journals	eBooks	Miscellaneous
Hydrological modeling AI	N/A	73	0	0
Machine Learning	N/A	211	0	0
Deep Learning	N/A	105	0	0
Flood Prediction	N/A	56	0	0
Rainfall prediction	N/A	174	0	2
AI rainfall prediction	N/A	56	0	0

#### 4. Data Extraction

In the second stage, we physically examined conference and journal paper titles (eBooks, publications prior to year 2020 and miscellaneous are omitted) of each one of the 695 (IEEE), 1167 (Springer) and 677 (Elsevier). Publications analyzed from rest of libraries as an introductory search, rejected any which are not related to our domain of research and also multiple abstract similarity data rejected if published on more than one database.



**Fig. 1: Literature database search results**

We filtered the unique data as per our domain of proposed research and finalized 127 documents with an immediate significance to our work. These were most significant and the information is about the ML/DL architectures, algorithms for flood prediction, reference architecture models etc. For immediate pilot study, we extracted 16 papers, as demonstrated in above Fig. 1.

##### 4.1. Quality assessment

We evaluated shortlisted papers for following aspects,

Q1: Is author cited high impact factor publications?

Q2: Is information provided by author is cited by other researchers?

Q3: Is existing analysis is used for real-time tasks?

The quality assessment answers are synchronized as follows:

Ans.1: Yes. But authors also referred print journals before year 2000.

Ans.2: Yes. Research citation record till 2019 is available with each e-article and many cross-references used in original cited article.

Ans.3: Most commonly authors used datasets and machine learning models.

##### 4.2. Data collection Strategy

The present data is gathered from high indexed journals, conferences and transactions for study purpose. The information is separated as:

- Title of research
- Author name of research paper
- Publication Date &Year, Volume, Issue, Page numbers, ISSN/DOI

##### 4.3. Data analysis

Data is organized in tabular format for further analysis (refer table 4). The tabulated data contains pilot search.

**Table 4: Final documents for review**

Ref. No.	Author	Year of publication
[1]	Gonzales-Inca, Carlos, et al.	2022
[2]	Dehghani, Reza, et al.	2022
[3]	Tan, Woon Yang, et al.	2022
[4]	Chen, Chen, et al.	2022
[5]	Motta, Marcel, et al.	2021
[6]	Lawal, et al.	2021
[7]	Lammers, Roderick, et al.	2021
[8]	Yan, Xingyu, et al.	2021
[9]	Wang, Yi, and Antonia Sebastian	2021
[10]	Kaur, Mandeep , et al.	2021
[11]	Moishin, Mohammed, et al.	2021
[12]	Speight, Linda J., et al.	2021
[13]	Samikwa, Eric, Thiemo Voigt, and Joakim Eriksson	2020
[14]	Wu, Wenyan, et al.	2020
[15]	Zanchetta, Andre DL, and Paulin Coulibaly	2020
[16]	Khan, Talha Ahmed, et al.	2020

## 5. Challenges Identified

- Data Quality & data availability is important to predict region based flood situations.
- Predictive model needs to be generalized so any dataset can be fit to prediction array of parameters. There is a need of standardization of weather parameters taking part in flood forecast.
- Interpretability of deep learning models needs to be enhanced in terms of graphical representation. So, any person can interpret the parameters.
- Computational resources important in flood prediction system and hence training on large datasets requires high-performance hardware. So, it is important to cut down the redundant requirement of hardware.
- Other core challenge when we refer existing studies for performance comparison, performance Metrics used by many authors are different; consequently comparison with existing systems becomes difficult. Ex. RMSE (Root Mean Square Error), MAE (Mean Absolute Error),  $R^2$  (Coefficient of Determination), Accuracy / Precision / Recall (for classification-based flood alerts) etc.

Further, notable limitation for performance analysis is the dataset used during the research.

- Satellite Data Retrieval - TRMM, GPM, MODIS
- Hydrological Station Data as- Rain gauge data, river discharge measurements
- Remote Sensing Data Retrieval- Soil moisture, terrain elevation

ML models have been widely used for predictive modeling in hydrology are identified as:

- **Linear Regression (LR):** Basic model, often used as a benchmark.

- **Support Vector Machines (SVM):** Performs well with non-linear relationships.
- **Random Forest (RF):** Good for capturing feature interactions and non-linearity.
- **K-Nearest Neighbors (KNN):** Effective but sensitive to data scale and noise.
- **Gradient Boosting Machines (GBM, XGBoost):** High accuracy, widely used in competitions and practical implementations.

DL models (as listed below) automatically extract features from raw data and have shown superior performance in time-series forecasting.

- **Artificial Neural Networks (ANN):** Widely used but require manual feature engineering.
- **Recurrent Neural Networks (RNN):** Handle sequential data; basic form of temporal modeling.
- **Long Short-Term Memory (LSTM):** Ideal for learning long-term dependencies in time-series flood data.
- **Convolutional Neural Networks (CNN):** Used for spatial rainfall pattern recognition.
- **Hybrid Models (e.g., CNN-LSTM):** Combine spatial and temporal modeling for better accuracy.

## 6. Conclusions

Different criteria can be used for the prediction of a flash flood scenario. Existing operational systems may take into account meteorological patterns known to precede extreme precipitation events and apply rainfall or runoff threshold comparisons over forecasted values to identify upcoming flood events. The choice of the “best” criteria to be adopted is mainly driven by the area covered by the forecasting team and by the available resources. Hence, distributed hydrological models based on mixed conceptual-physical representations of the rainfall-runoff/routing processes are being preferred to their physically-based counterparts due to the reduced number of parameters to be calibrated. IoT, cloud computing, data based and artificial intelligence models are being preferred to improve the quick forecast system. Hence, there is wide scope for DL/ML systems to collaborate with standardized dataset as well as IoT systems.

## References:

- [1] Gonzales-Inca, Carlos, et al. "Geospatial artificial intelligence (GeoAI) in the integrated hydrological and fluvial systems modeling: Review of current applications and trends." *Water* 14.14 (2022): 2211.
- [2] Dehghani, Reza, Hamidreza Babaali, and Nejat Zeydalinejad. "Evaluation of statistical models and modern hybrid artificial intelligence in the simulation of precipitation runoff process." *Sustainable Water Resources Management* 8.5 (2022): 154.
- [3] Tan, Woon Yang, et al. "State-of-the-art development of two-waves artificial intelligence modeling techniques for river streamflow forecasting." *Archives of Computational Methods in Engineering* 29.7 (2022): 5185-5211.
- [4] Chen, Chen, et al. "A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China." *Journal of Hydrology* 607 (2022): 127535.
- [5] Motta, Marcel, Miguel de Castro Neto, and Pedro Sarmiento. "A mixed approach for urban flood prediction using Machine Learning and GIS." *International journal of disaster risk reduction* 56 (2021): 102154.
- [6] Lawal, Zaharaddeen Karami, Hayati Yassin, and Rufai Yusuf Zakari. "Flood prediction using machine learning models: a case study of Kebbi state Nigeria." *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. IEEE, 2021.
- [7] Lammers, Roderick, et al. "Prediction models for urban flood evolution for satellite remote sensing." *Journal of hydrology* 603 (2021): 127175.

- [8] Yan, Xingyu, et al. "A rapid prediction model of urban flood inundation in a high-risk area coupling machine learning and numerical simulation approaches." *International Journal of Disaster Risk Science* 12.6 (2021): 903-918.
- [9] Wang, Yi, and Antonia Sebastian. "Community flood vulnerability and risk assessment: An empirical predictive modeling approach." *Journal of Flood Risk Management* 14.3 (2021): e12739.
- [10] Kaur, Mandeep, Pankaj Deep Kaur, and Sandeep Kumar Sood. "Energy efficient IoT-based cloud framework for early flood prediction." *Natural Hazards* 109.3 (2021): 2053-2076.
- [11] Moishin, Mohammed, et al. "Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm." *Ieee Access* 9 (2021): 50982-50993.
- [12] Speight, Linda J., et al. "Operational and emerging capabilities for surface water flood forecasting." *Wiley Interdisciplinary Reviews: Water* 8.3 (2021): e1517.
- [13] Samikwa, Eric, Thiemo Voigt, and Joakim Eriksson. "Flood prediction using IoT and artificial neural networks with edge computing." *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. IEEE, 2020.
- [14] Wu, Wenyan, et al. "Ensemble flood forecasting: Current status and future opportunities." *Wiley Interdisciplinary Reviews: Water* 7.3 (2020): e1432.
- [15] Zanchetta, Andre DL, and Paulin Coulibaly. "Recent advances in real-time pluvial flash flood forecasting." *Water* 12.2 (2020): 570.
- [16] Khan, Talha Ahmed, et al. "Investigation of flash floods on early basis: A factual comprehensive review." *IEEE Access* 8 (2020): 19364-19380.