Empowering Communities Through Resilient River Management

Gresha Bhatia

gresha.bhatia@ves.ac.in
Department of Computer Engineering, Vivekanand Education Society's Institute of Technology
Mumbai, India - 400074

KEYWORDS

Flood management systems, Machine learning algorithms, prediction of floods, overflow, safe and evaluation measures

ABSTRACT

This paper focuses on empowering the society with various rescue strategies in the event of a natural disaster namely floods. This paper deals with applying the state of the art technologies such as machine learning in order to combat floods. The key aspects of a flood management system and the existing systems were then identified. The proposed system was then developed. Statistical, hydrological and Machine learning algorithms such as Random forest, support vector machines, Naïve Bayes, decision tree algorithms were further applied on the data parameters captured. to predict the overflow probability in different regions of the state. A comparison was then undertaken to determine the most suited algorithm to perform better river management. The paper then concluded highlighting the need for applying machine learning techniques for flood management.

I. INTRODUCTION

A Flood Prediction System forms a critical component of disaster management. It aims to forecast the likelihood, timing, and severity of flooding events. It leverages advanced technologies, hydrological data, and mathematical models to provide accurate and timely predictions. These systems empower authorities and communities to prepare effectively, reducing loss of life and property. In recent times, however, the frequency and severity of flood-related disasters have increased significantly, posing substantial challenges globally as represented in figure 1.

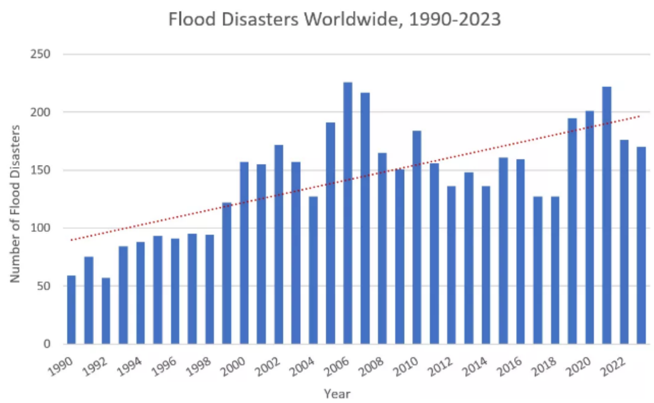

Figure 1:- Flood disasters worldwide

Source:-

www.statista.com/statistics/1339730/number-of-flood-disasters-worldwide/

These catastrophic occurrences often result in loss of life, displacement of communities, and extensive damage to infrastructure and livelihoods.

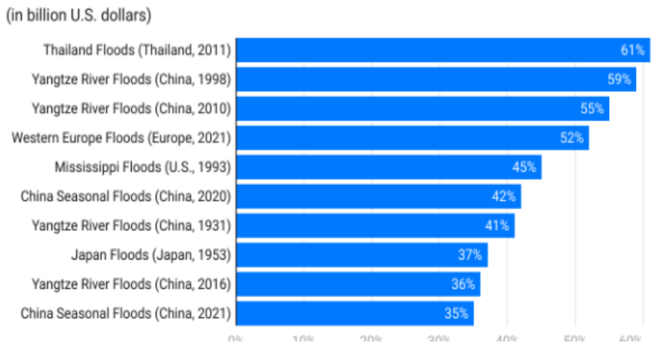

Figure 2 : Most expensive flood disasters worldwide

Source : www.coolest-gadgets.com/statistics-about-flooding-and-water-damage-in-the-usa/

Recent findings predict a rise in both the frequency and intensity of local flooding, as extreme precipitation events become more common due to climate change. Flooding, the most common weather-related natural disaster, has intensified in recent decades, leading to profound losses worldwide as evident from figure 2. Despite advancements in meteorological and hydrological sciences, the importance of accurate flood prediction and prompt disaster response cannot be overstated in mitigating the impact of these events. Flood prediction systems are thus essential for mitigating the impacts of climate change, which has intensified flooding events globally. By combining advanced machine learning techniques, historical meteorological data, and analysis of terrain, these systems can help to build resilient and informed societies. This paper focus on the development of predictive flood management model. The flow of the paper is as follows. Section II highlights the key features of flood management systems. Section III identifies the related work done in this domain. The need of developing the proposed system is elaborated in Section IV. Machine learning algorithms are further explored in section V . The proposed system and its respective process design is mentioned in section VI and VII respectively. The results obtained from the proposed system is elaborated in section VIII. Section IX provide the conclusion drawn .

II. KEY FEATURES OF FLOOD MANAGEMENT SYSTEMS

This section elaborates on the critical aspects that make up the flood management system. These include:-

1. **Data Collection and Monitoring:**

It focuses on continuously monitoring the rainfall, river levels, soil saturation, and weather conditions using sensors, satellites, and weather stations. Trend analysis can further be performed by integrating the historical flood related data.

2. **Hydrological and Meteorological Models:**

In this various models including climate models can be incorporated to simulate river flows, rainfall-runoff processes, inundate patterns and predict weather patterns contributing to floods.

3. **Real-Time Forecasting:**

Machine learning algorithms are employed that work closely with prediction models. These models are supposed to operate on real time data feeds so that near-instantaneous updates and improvement in forecast accuracy can be obtained.

4. **Early Warning Systems (EWS):**

Risk maps and dashboards showing potential inundation areas along with automated alerts and notifications can be sent to authorities.

5. **Decision Support Systems:**

Scenario-based simulations Tools can be used to to assist in planning evacuation routes, resource allocation, and response coordination.

6. **Community and Stakeholder Engagement:**

Information to the community at large and collaboration with response teams for comprehensive preparedness can be provided in accessible formats.

III. LITERATURE SURVEY

This section comprises of a comprehensive review of the work done in the domain of flood management existing along with the articles refereed to . The details are as mentioned below :-

Research papers referred:

A number of research papers were referred to as specified below. The authors expressed that Employ a random forest model had been employed to pinpoint high-risk flood zones.[1]. Utilizing advanced GIS models to generate flood susceptibility maps for Seoul, incorporating critical factors such as river proximity and topography, alongside validation of predictive accuracy, were the key findings mentioned the authors [2]. Through paper [3], the authors presented an overview of the various machine learning algorithms that can be incorporated in flood predictions. The study in [4] focused on flood hazards in the lower Narmada basin, using flood frequency analysis and hydrodynamic modeling to predict peak floods and produce inundation maps. It applied Log-Pearson Type III distribution and Gumbel's method for lower and higher return periods respectively. This provided critical insights for flood mitigation planning. In paper [5] the authors elaborated on the forecasting systems mentioned by the Central Water Commission that relied on statistical methods and real-time data collection from field stations via wireless, telephone/mobile, and satellite telemetry. In-house developed models utilized rainfall data from various sources for predictions, disseminated through dedicated websites and social media. The "Flood Management Programme (FMP)," as outlined in [6], was introduced by the Ministry during the XI Plan as a State Sector Scheme and continued into the XII PlanIn [7], the authors improved flood forecasting at the Sardar Sarovar Dam by employing clustering techniques and Thiessen polygons to identify key rain gauge stations. The study concludes that the current gauge network is adequate for forecasting and highlights the critical role of selecting appropriate stations for accurate predictions in flood-prone regions.

IV. EXISTING SYSTEMS

Current systems fail to address dynamic factors such as :- 1. Shifting land use patterns and infrastructure development. 2. Assumptions in input data and uncertainties in future climate scenarios also contribute to inaccuracies. 3. Inadequate parameters further reduce the precision of predictions. 4. The absence of sufficient and reliable historical flood data undermines the system's overall effectiveness. Simulated flood maps may also have uncertainties due to reliance on simplified models and assumptions. 5. Government-operated models depend on dam/reservoir data being shared with the CWC through WIMS or other communication channels like email, SMS, or phone, which results in delays while waiting for data processing and action decisions. Such delays could be catastrophic in emergencies 6. Many projects focus predominantly on flood management and hydrological aspects, often neglecting the role of predictive modeling. 7.Limited data records and overly simplistic assumptions about rainfall-runoff relationships and land use further hinder the accuracy and generalizability of these models.

V. PROPOSED SYSTEM

The objective behind the development of the proposed system is to apply various machine learning algorithms towards prediction of floods. This would further be employed by emergency and rescue teams for near real time mitigation management. As machine learning models enables the processing of large volumes of data that is input, it aids in extracting insights and identify patterns [8].

1.Parameters used in the proposed model:-

As machine learning algorithms produce better and finer results if more data is input to it, various data inputs are fed to it namely:

a.Design gross Storage:- It defines the total volume or capacity available in a storage unit

b. Present Gross Storage:-

It defines the total capacity of a storage system including the useful and dead storage

c. Outflow River:-

It forms a watercourse that carries water from a reservoir, lake, or dam downstream into a natural channel or another storage system

d. Cummulative Rainfall:-

Cumulative rainfall refers to the total amount of rainfall that has accumulated over a specific period, typically measured in millimeters (mm) or inches

e. Gate-Position-Nos:-

Gate-Position-Nos refers to the identification or numbering of gates used to control water flow. These gates regulate the release, diversion, or containment of water in a structured and monitored manner

f.Scheme:-

It is used to address challenges, optimize resources, and ensure systematic progre

g. Present-Water Level(m):-

It specifies the current height or depth of water in a reservoir, lake, river, canal, or other water body. It is a critical parameter for monitoring and managing water resources,

h. Inflow:-

It refers to the volume of water entering a reservoir, river, lake, canal, or any other water storage over a specific period.

Using the above mentioned data, a relationship is established between the level of the reservoir and the gate opening of the reservoir. This relationship along with various other input parameters can be used to train the machine learning algorithm such as random Forest Regression Model.

2.Flow of the system :-

The entire system comprises of the following steps

a. Data Preprocessing

Initially in this step, the entire dataset comprising of the parameters mentioned above are loaded forming the dataset for operations. Once the data is obtained ,Irrelevant features are eliminated and the missing values are corresponding filled with the mean of respective columns. Further, Categorical variables are encoded into numerical format as depicted in figure 3. The flow of the system is as represented in figure 4.

Outflow canal is an engineered channel that carries water away from a reservoir, dam, or other water storage systems to designated areas such as rivers, irrigation fields, industrial zones.

i.Type of Gate:-

They control water flow, manage reservoir levels, and regulate discharge to downstream areas

Scheme	Scheme	Design C	FRL (m)	Rule Lev	Present	Present	Percenta	Inflow (C	Outflow F	Outflow C	Cumm. R	Type		Gate Pos	Opening(m)
36	Ukai	7414.3	105.16	105.16	105.01	7328.6	98.85	18261	0	800	1431	G		0	0
42	Damang	524.05	73.05	73.05	73.05	524.05	100	4205	4305	75	2465.6	G		2	0.3
27	Watrak	158.2	136.25	136.25	133.87	103.76	65.59	0	0	0	842	G		0	0
28	Guhai	68.75	173	173	170.25	38.39	55.85	82	0	0	772	G		0	0
29	Mazam	43.86	157.1	157.1	154.28	20.89	47.64	25	0	0	875	G		0	0
30	Hathmat	152.93	180.75	0	178.84	74.16	48.5	0	0	0	797	UG		0	0
32	Javanpu	2.5	91	91	91	2.5	100	180	180	0	755	G		1	0.1
33	Harnav-I	21.87	332	332	331.5	20.52	94.72	0	0	0	728	G		0	0
34	Meshwo	53.13	214.59	0	210.36	28.82	54.25	80	0	0	740	UG		0	0
12	Wanakbo	41.88	87.3	0	88.12	46.61	100	16036	13586	2750	1006.6	UG		0	0
14	Panam	578.19	127.41	127.41	127.41	578.18	100	2301	2852	200	939	G		1	0.6
16	Hadaf	22.09	166.2	166.2	166.2	22.09	99.98	690	681	0	700	G		1	0.15
17	Kadana	1249.3	127.71	127.71	127.66	1244.2	99.59	12119	5100	700	821	G		0	0
6	Karjan	538.75	115.25	115.25	114.24	523.04	97.08	4207	1461	0	878	G		1	0.2
40	Sukhi	173.01	147.82	147.82	147.78	174.3	100	521.95	521.95	0	1172	G		1	0.15
3	Mukteshv	31.46	201.65	201.65	199.8	23.26	73.93	100	0	0	917	G		0	0
4	Dantiwac	397.12	184.1	184.1	184.1	393.62	99.12	395	395	2	825	G		1	0.3
5	Sipu	161.43	186.43	186.24	180.6	53.22	32.97	0	0	0	480	G		0	0
13	Dharoi	813.14	189.59	189.28	189.33	785.52	96.65	2196	1506	650	818.4	G		1	0.3
65	Khodiyar	29.94	202.68	202.68	201.25	22.74	75.97	130	0	0	820	G		0	0
76	Shetrunj	346.48	55.53	55.53	55.53	346.48	100	0	0	0	540	G		0	0
94	Und-I	69.05	98	98	98	69.05	100	321	321	0	844	G		1	0.15
146	Bhadar	188.14	107.9	107.9	107.35	164.35	87.36	35	0	0	765	G		0	0
149	Bhadar -	49	53.1	53.1	53.1	49	100	0	0	0	825	G		0	0
150	Machchl	87.9	57.3	57.3	56.12	67.21	76.46	430	0	0	589	G		0	0
156	Machchl	68.95	135.33	0	135.02	61.17	88.71	0	0	0	825	UG		0	0

Figure 3:- Data parameters considered for the proposed system

b. Training the model:-

The dataset is further split into the training and testing set on th basis of 80:20 rule where 80% of the data is used for training and 20% for testing. A machine learning model such as the Random Forest Classifier: A Random Forest Classifier or a predictive model is then applied on the given input.

c. Graphical User Interface (GUI):-

The graphical interface was further created using Tkinter Window and the Flask. This enabled inputting the various parameters and obtains predictions.

d. Execution of the proposed algorithm

Based on the input provided through the above mentioned steps and the trained Random Forest

Model, the percentage storage of the specified reservoir is predicted to be within safe limits or if there utilized for classification and regression.
..

VI. ALGORITHMS INCORPORATED

Various machine learning algorithms were identified and executed to get better water resource management and thereby help in informed , real time decision making . These algorithms. The algorithms ranged from classification to prediction methods namely :- **i. Random Forest Algorithm :-**
Through this algorithm, multiple decision trees were trained on diverse subsets.

Advantages:-

i.It resulted in obtaining the predictions through majority voting
ii. It gave enhanced predictive accuracy
iii. It prevented the reliance on a single decision tree.
iv. It addressed overfitting concerns.

The mathematical formula utilized for Random Forest was :-
nij = WjCj-left(j) Cleft(j) - Wright(j) Cright(j)

This algorithm was then tailored to integrate real-time data on water levels and hydrological factors from various schemes of the Narmada river.

ii. Support Vector Machine (SVM) algorithm:-

This algorithm was employed to transform data, establish optimal boundaries and thereby distinguish between different outputs[11]

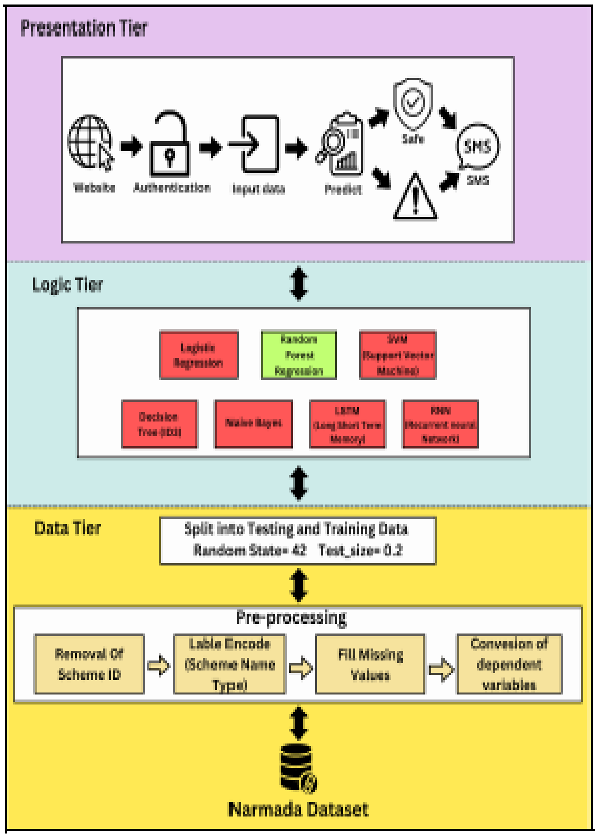

Fig. 4 Flow of Proposed System

iii. Logistic Regression
Logistic regression, a supervised learning algorithm
was used for categorical and labeled target variables classification. It was applied to provide a binary predicted classification into a 0 or 1 category, based on a threshold of 0.5. The formula used for the same is as shown by $y = mx + b$.

iv. Decision Tree algorithm:-
It formed a supervised machine learning algorithm for classification purposes. It operates on a tree-like structure, where each node represented a test on a particular attribute while each branch determined the outcome from that test. The leaf node indicated the class label associated with the path it led to. The value of the node in decision-making decreased as we moved down the tree. Various techniques were incorporated for carrying out the decision namely:-

1. CART (Classification and Regression Tree)
   It used the Gini index as a metric and was represented as

$$\text{gini}_{A}(D) = \frac{|D1|}{D}\,\text{gini}(D_1) + \frac{|D2|}{D}\,\text{gini}(D_2)$$

2. ID3 (Iterative Dichotomiser) algorithm :-
   This algorithm uses Entropy Function and Information Gain as its metrics.
   Formula for Information Gain:

$$I(P_i, N_i) = \frac{-p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

   Formula for Entropy:

$$\Sigma \left(\frac{Pi+Ni}{p+n}\right) I(P_i, N_i)$$

V. . Naive Bayes algorithm:-
It forms a probabilistic based classification machine learning algorithm. Due to its ease of implementation, faster computations, quick training and predictions, it is applied in scenarios having high-dimensional inputs. The mathematical formula incorporated in this paper is as follows:

$$P(A|B)=P(B|A).P(A)/P(B)$$

Where,
A and B are two events
P(A/B) is the conditional probability that event A occurs, given that event B has occurred.
P(B|A) is the conditional probability that event B
occurs, given that event A has occurred.
P(A) and P(B) denote individual probabilities of A and B respectively.

6. Long Short-Term Memory (LSTM) Network:-
It introduces a memory cell that is capable of capturing prolonged dependencies within sequential data and retaining information over extended period of time.
It provides a base for time series forecasting[9] [10].
The activation function used in LSTM is as follows :-
Activation at time *t* :

$$h_t^j = (1- z_t^j)h_{t-1}^j + z_t^j \tilde{h}_t^j$$

Update gate:

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j$$

Candidate activation:

$$\tilde{h}_t^j = tanh(Wx_t + U(rt \otimes h_{t-1}))^j$$

Reset gate:

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j$$

Where:
Activation at time t (h_t^j)

7. Prediction Using Recurrent Neural Network (RNN)
It is used to process sequential data as it incorporates feedback loops. This allows it to be considered for the current as well as the input knowledge obtained from past inputs. As it assigns weights to both the current as well as previous input, it can iteratively adjust the weights through gradient descent and backward propagation [3]. This helps in better decision making as compared to the feed forward Neural Networks. The mathematical formulations for RNN is as follows:

$$o^t = f(h^t, \theta)$$
$$h^t = k(h^{t-1}, x^t, \theta)$$

Where, o^t is the output produced at time t,
h^t is the state of hidden layers at time t,

x^t is the input given at time t,
θ indicates the weights and biases of that particular network

VII. Results and Discussions

The results obtained when the various algorithms were implemented are as mentioned below.

a. Classification Results of various algorithms:-

Figure 5 - 11, represent the classification obtained when the algorithms such as SVM , Logistic regression, Random Forest Regression, Decision Tree, Naive Bayes classification algorithm , LSTM and RNN models are respectively incorporated

Accuracy: 0.9411764705882353
Classification Report:

	precision	recall	f1-score	support
0	1.00	0.83	0.91	6
1	0.92	1.00	0.96	11
accuracy			0.94	17
macro avg	0.96	0.92	0.93	17
weighted avg	0.95	0.94	0.94	17

Fig 5:- Classification obtained through SVM

Accuracy: 0.7647058823529411
Classification Report:

	precision	recall	f1-score	support
0	0.60	1.00	0.75	6
1	1.00	0.64	0.78	11
accuracy			0.76	17
macro avg	0.80	0.82	0.76	17
weighted avg	0.86	0.76	0.77	17

Fig. 6 Classification obtained through Logistic Regression

Accuracy: 0.9411764705882353
Classification Report:

	precision	recall	f1-score	support
0	1.00	0.83	0.91	6
1	0.92	1.00	0.96	11
accuracy			0.94	17
macro avg	0.96	0.92	0.93	17
weighted avg	0.95	0.94	0.94	17

Fig 7 : Classification obtained by Random Forest Regression

Accuracy: 0.9411764705882353
Classification Report:

	precision	recall	f1-score	support
0	1.00	0.83	0.91	6
1	0.92	1.00	0.96	11
accuracy			0.94	17
macro avg	0.96	0.92	0.93	17
weighted avg	0.95	0.94	0.94	17

Fig. 8 Classification obtained by Decision Tree

Accuracy: 0.9411764705882353
Classification Report:

	precision	recall	f1-score	support
0	0.86	1.00	0.92	6
1	1.00	0.91	0.95	11
accuracy			0.94	17
macro avg	0.93	0.95	0.94	17
weighted avg	0.95	0.94	0.94	17

Fig. 9 Classification obtained by Naive Bayes algorithm

```
Accuracy: 0.8823529411764706
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.83      0.83         6
           1       0.91      0.91      0.91        11

    accuracy                           0.88        17
   macro avg       0.87      0.87      0.87        17
weighted avg       0.88      0.88      0.88        17
```

Fig. 10 Classification obtained by LSTM network

```
Accuracy: 0.7058823529411765
Classification Report:
              precision    recall  f1-score   support

           0       0.67      0.33      0.44         6
           1       0.71      0.91      0.80        11

    accuracy                           0.71        17
   macro avg       0.69      0.62      0.62        17
weighted avg       0.70      0.71      0.67        17
```

Fig. 11. Classification report of Recurrent Neural Network

b. Evaluation metrics of algorithms implemented:-

Evaluation metrics are determined through the various measures of accuracy, precision, recall, F1 score. The algorithms incorporated were compared on the above mentioned metrics. The results for the same are presented below through table 1.

Sr. No	Algorithm	Accuracy	Precision	Recall	F1 Score
1	SVM	0.9411	0.92	1.00	0.96
2	Logistic	0.7647	1.00	0.64	0.78
3	Random Forest	0.9411	0.92	1.00	0.96
4	Decision Tree	0.9411	0.92	1.00	0.96
5	Naïve Bayes	0.9411	1.00	0.91	0.95
6	LSTM	0.8823	0.91	0.91	0.91
7	RNN	0.7058	0.71	0.91	0.80

Table 1: Comparison of various machine learning algorithms

From table 1, it is observed that algorithms SVM, random forest, decision tree as well as LSTM give accurate results followed by LSTM, Logistic regression and RNN respectively.
As compared to other algorithms, Logistic and Naïve Bayes provide precise results.
Looking at these two characteristics, it can be said that Naïve Bayes performs better in comparison with the other algorithms considered.

VIII. CONCLUSION

The paper elaborates on the measures that need to be taken in order to provide a better resiliency towards flood scenario. Machine learning algorithms were then elaborated upon. This was done to indicate that the current state of the art technologies can help in better rescue operations and provide for better decision making. The proposed reservoir flood warning system further used advanced flood prediction models and real-time data. Their integration could help in issuing timely alerts and warning for rescue operations. The system aims to enhance disaster preparedness and minimize flood risks.

REFERENCES

[1] Zhaoli Wang , Chengguang Lai , Xiaohong Chen , Bing Yang , Shiwei Zhao , Xiaoyan Bai, Flood hazard risk assessment model based on random forest , Year of Publication: August 2015.
[2] Sunmin Lee, Jeong-Cheol Kima, Hyung-Sup Junga,
Moung Jin Leecand Saro Lee, Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea , Year of Publication: 10 Apr 2017.
[3] Chinmayee Kinage, Sejal Mandora, Abhishek
Kalgutkar, Sunita Sahu , Amruta Parab, Performance
Evaluation of Different Machine Learning Based
Algorithms for Flood Prediction and Model for Real Time Flood Prediction, Year of Publication: 2019, 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA).
[4] Nikunj K. Mangukiya; Darshan J. Mehta; Raj Jariwala, Flood frequency analysis and inundation mapping for lower Narmada basin, India, Year of Publication: 1 February 2022.
[5] Minister of State for Jal Shakti, Shri Bishweswar Tudu, Flood Forecasting and Early Warning System , Publication Date : 08 AUG 2022 6:02PM by PIB Delhi
[6] Press Information Bureau, Government of India,
Ministry of Water Resources, River Development and
Ganga Rejuvenation, Flood Control Schemes by Indian
Government, Year of Publication: January 10, 2023
[7] Darshan Mehta , Jay Dhabuwala , Sanjaykumar M.
Yadav , Vijendra Kumar , Hazi M. Azamathulla ,
Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modeling, Year of Publication: December 2023
[8] A Yovan Felix , T Sasipraba , Flood Detection Using
Gradient Boost Machine Learning Approach , Year of
Publication: December 2019 , Conference: 2019
International Conference on Computational Intelligence
and Knowledge Economy (ICCIKE)
[9] Jiaojiao Hu, Xiaofeng Wang, Ying Zhang, Depeng
Zhang, Meng Zhang, Jianru Xue , Time Series Prediction Method Based on Variant LSTM Recurrent NeuralNetwork ,Neural Processing Letters, 2020
[10] Pin Zhang , Zhen-Yu Yin , State of the-Art Review of Machine Learning Applications in Constitutive Modeling of Soils, Archives of Computational Methods in Engineering, 2021
[11] Yutaro Fuse, Yoshitaka Nagashima, Hiroshi
Nishiwaki, Fumiharu Ohka et al., Development of machine learning models for predicting unfavorable functional outcomes in patients with chronic subdural hematomas ,Research Square Platform LLC, 2022