Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989 –9572

DOI: 10.47750/jett.2023.14.02.059

FLOOD FORECASTING USING MACHINE LEARNING

[1]*Mrs.M.Eramma*
[2]*Kamarathi Swetha*
[3]*Ganapa Sudha*
[4]*Jinde Jayasree Bai*
[5]*Jagatapu Sulochana*

Journal for Educators,Teachers and Trainers, Vol.14 (2)

https://jett.labosfor.com/

Date of reception: 18 Jan 2023

Date of revision: 09 Feb 2023

Date of acceptance: 21 Mar 2023

Mrs.M.Eramm, Kamarathi Swetha, Ganapa Sudha, Jinde Jayasree Bai, Jagatapu Sulochana (2023). **FLOOD FORECASTING USING MACHINE LEARNING**. Journal for Educators, Teachers and Trainers,Vol. 14(2). 634-644.

Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 14(2)

ISSN 1989 –9572

https://jett.labosfor.com/

FLOOD FORECASTING USING MACHINE LEARNING

[1]Mrs.M.Eramma,[2]Kamarathi Swetha,[3]Ganapa Sudha,[4]Jinde Jayasree Bai,[5]Jagatapu Sulochana,
[1]Assistant Professor,[2345]Students
Department Of CSE
St. Johns College of Engineering & Technology, Errakota, Yemmiganur

ABSTRACT:
A flood is the inundation of a significant volume of water that exceeds the capacity of a land area, resulting in overflow. The flood prediction (FF) system issues warnings based on the water level or discharge rates via hydraulic infrastructure. Flood forecasting (FF) enhances the capacity and progress in hydrology to reduce risks by using machine learning techniques, namely artificial neural networks (ANN). The use of machine learning algorithms (MLAs) in flood forecasting aims to enhance the system's capacity to predict and reduce flood risks in response to climate change. This study is being conducted to anticipate floods in the Upper Wardha project within the Wardha river basin. Flood forecasting (FF) involves the use of real-time estimate to determine the likelihood of a flood occurring. By using the predicted inflow, the pace at which water enters a reservoir, the timing of operations such as opening and shutting gates may be determined in real-time using artificial neural networks (ANN).

I. INTRODUCTION

Flood Forecasting can be defined as a process of estimation of time and duration based on topographical characteristics of any river basin which reduced the hazards to human life and environment also. Flood Forecasting technique challenge to predict occurrence and magnitude with time of flash flood. Flood happened due to continue precipitation with respective time. Ordinary rainfall also contributes to transform with time into deadly flood. Flood forecasting techniques play a vital and important role to mitigate the hazards for non-structural structures with cost- effective management. Flood forecasting stations cover the network of flood prone area to give flood warning to administration. Forecasting inflow also used for the operation of hydraulic structures such as dam on which opening and closing of gates on spillways. Flood forecasting techniques and Flood warning system are required different type of architectures of flood. Flood may be reduced by constructing structures such as dams, weirs, dykes but cannot eliminate the risk. Flood forecasting techniques able to mitigate the hazards for population and environment in real time with an early warning [1]. Flood forecasting has been approach through rainfall – runoff and flood routing model. Flood forecast predict inflow at selected location with HFL value at selected locations of river with time depends on watershed or catchment area. Later, downstream side predicts the flood limited with travel time with an assessment of uncertainties to properly support the decision-makers activities [9]. Flood forecasting using machine learning algorithm (MLA) method understand to learn and improve system scale to mitigate flood hazards according to the climate change with the help of AI. Flood forecasting used for creating the machine learning algorithm with past and real records of flood with real time data using rain gauges for different coming back time. The dataset sources are rainfall-runoff, water levels using automatic rain gauges with satellites technology, infiltration rate etc

Problem Statement:

Flood forecasting is a critical aspect of disaster management, providing valuable insights and early warnings to mitigate the impact of flooding on communities and infrastructure. Traditional flood forecasting methods often rely on simplistic models and historical data, struggling to capture the complexities of dynamic environmental factors that contribute to flooding. Additionally, the increasing frequency and intensity of extreme weather events, exacerbated by climate change, pose new challenges for accurate and timely flood predictions.

The limitations of current flood forecasting systems include the inability to adapt to rapidly changing weather patterns, inadequate spatial resolution, and challenges in integrating diverse data sources. As a result, there is a pressing need for advanced and adaptive forecasting approaches that harness the power of machine learning to improve accuracy, incorporate real-time data, and enhance the lead time for flood warnings.

The existing problem lies in the inefficiency of conventional flood forecasting methods to provide precise and timely predictions, especially in the face of changing climate patterns and extreme weather events. Developing machine learning-based flood forecasting models is crucial to address these limitations and advance the capabilities of flood prediction systems for better disaster preparedness and response.

Objectives :

1. Data Integration and Preprocessing:
   - Aggregate and preprocess diverse data sources, including weather patterns, river discharge, soil moisture, and historical flood data, for input into machine learning models.
2. Feature Selection and Extraction:
   - Identify and extract relevant features from the integrated datasets, employing techniques that enhance the model's ability to capture patterns indicative of flood events.
3. Machine Learning Model Selection:
   - Evaluate and select appropriate machine learning models, such as regression models, decision trees, ensemble methods, or deep learning architectures, based on the characteristics of the flood forecasting problem and the available data.
4. Real-Time Data Integration:
   - Develop mechanisms for real-time integration of weather and hydrological data, enabling the model to adapt to changing conditions and provide up-to-date flood predictions.
5. Spatially Resolved Forecasting:

- Enhance the spatial resolution of flood predictions by employing machine learning techniques that consider the local topography and geographical features, ensuring more accurate and localized forecasts.

6. Uncertainty Estimation:
   - Implement methods to estimate and communicate uncertainties associated with flood predictions, providing decision-makers and communities with a clearer understanding of the reliability of forecasted outcomes.
7. Seasonal and Climate Variability Considerations:
   - Incorporate features and patterns related to seasonal and climate variability, ensuring the machine learning models account for long-term trends and changes in climate that may influence flood occurrences.

II. LITERATURE SURVEY

Pappenberger, F.; Cloke, H.L.; Parker, D.J.; Wetterhall, F.; Richardson, D.S.; Thielen, J.The monetary benefit of early flood warnings in Europe. Environ. Sci. Policy 2015, 51, 278–291.

Effective disaster risk management relies on science-based solutions to close the gap between prevention and preparedness measures. The consultation on the United Nations post-2015 framework for disaster risk reduction highlights the need for cross-border early warning systems to strengthen the preparedness phases of disaster risk management, in order to save lives and property and reduce the overall impact of severe events. Continental and global scale flood forecasting systems provide vital early flood warning information to national and international civil protection authorities, who can use this information to make decisions on how to prepare for upcoming floods. Here the potential monetary benefits of early flood warnings are estimated based on the forecasts of the continental-scale European Flood Awareness System (EFAS) using existing flood damage cost information and calculations of potential avoided flood damages. The benefits are of the order of 400 Euro for every 1 Euro invested. A sensitivity analysis is performed in order to test the uncertainty in the method and develop an envelope of potential monetary benefits of EFAS warnings. The results provide clear evidence that there is likely a substantial monetary benefit in this cross-border continental-scale flood early warning system. This supports the wider drive to implement early warning systems at the continental or global scale to improve our resilience to natural hazards.

Krzysztofowicz, R. Bayesian system for probabilistic river stage forecasting. J. Hydrol. 2002, 268, 16–40.

The purpose of this analytic-numerical Bayesian forecasting system (BFS) is to produce a short-term probabilistic river stage forecast based on a probabilistic quantitative precipitation forecast as an input and a deterministic hydrologic model (of any complexity) as a means of simulating the response of a headwater basin to precipitation. The BFS has three structural components: the precipitation uncertainty processor, the hydrologic uncertainty processor, and the integrator. A series of articles described the Bayesian forecasting theory and detailed each component of this particular BFS. This article presents a synthesis: the total system, operational expressions, estimation procedures, numerical algorithms, a complete example, and all design requirements, modeling assumptions, and operational attributes.

Clark, M.P.; Slater, A.G. Probabilistic quantitative precipitation estimation in complex terrain. J. Hydrometeorol. 2006, 7, 3–22.

This paper describes a flexible method to generate ensemble gridded fields of precipitation in complex terrain. The method is based on locally weighted regression, in which spatial attributes from station locations are used as explanatory variables to predict spatial variability in precipitation. For each time step, regression models are used to estimate the conditional cumulative distribution function (cdf) of precipitation at each grid cell (conditional on daily precipitation totals from a sparse station network), and ensembles are generated by using realizations from correlated random fields to extract values from the gridded precipitation cdfs. Daily high-resolution precipitation ensembles are generated for a 300 km × 300 km section of western Colorado (dx = 2 km) for the period 1980–2003. The ensemble precipitation grids reproduce the climatological precipitation gradients and observed spatial correlation structure. Probabilistic verification shows that the precipitation estimates are reliable, in the sense that there is close agreement between the frequency of occurrence of specific precipitation events in different probability categories and the probability that is estimated from the ensemble. The probabilistic estimates have good discrimination in the sense that the estimated probabilities differ significantly between cases when specific precipitation events occur and when they do not. The method may be improved by merging the gauge-based precipitation ensembles with remotely sensed precipitation estimates from ground-based radar and satellites, or with precipitation and wind fields from numerical weather prediction models. The stochastic modeling framework developed in this study is flexible and can easily accommodate additional modifications and improvements.

Vrugt, J.A.; Robinson, B.A. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources.2007, 43.

Predictive uncertainty analysis in hydrologic modeling has become an active area of research, the goal being to generate meaningful error bounds on model predictions. State-space filtering methods, such as the ensemble Kalman filter (EnKF), have shown the most flexibility to integrate all sources of uncertainty. However, predictive uncertainty analyses are typically carried out using a single conceptual mathematical model of the hydrologic system, rejecting a priori valid alternative plausible models and possibly underestimating uncertainty in the model itself. Methods based on Bayesian model averaging (BMA) have also been proposed in the statistical and meteorological literature as a means to account explicitly for conceptual model uncertainty. The present study compares the performance and applicability of the EnKF and BMA for probabilistic ensemble streamflow forecasting, an application for which a robust comparison of the predictive skills of these approaches can be conducted. The results suggest that for the watershed under consideration, BMA cannot achieve a performance matching that of the EnKF method.

Ebtehaj, M.; Moradkhani, H.; Gupta, H.V. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap re-sampling. Water Resour. Res. 2010, 46.

Modeling of natural systems typically involves conceptualization and parameterization to simplify the representations of the underlying process. Objective methods for estimation of the model parameters then require optimization of a cost function, representing a measure of distance between the observations and the corresponding model predictions, typically by calibration in a static batch mode and/or via some dynamic recursive optimization approach. Recently, there has been a focus on the development of parameter estimation methods that appropriately account for different sources of uncertainty. In this context, we introduce an approach to sample the optimal parameter space that uses nonparametric block bootstrapping coupled with global optimization. We demonstrate the applicability of this procedure via a case study, in which we estimate the parameter uncertainty resulting from uncertainty in the forcing data and evaluate its impacts on the resulting streamflow simulations.

III. SYSTEM ANALYSIS

EXISTING SYSTEM :

The present study carried out on Wardha River basin .The Upper Wardha project is one of the major irrigation project in Vidharbha region of Maharashtra state. This project is across Wardha River near village SimbhoraTalukaMorshi of Amravati district. The Upper Wardha project consists of earthen dam on a both flanks with a centrally located gated spillway and canal on left and right flanks. The Upper Wardha dam has total grass irrigation (GI) of 11690 (ha) with gross capacity of 786 MCM in Godavari River Basin with annual rainfall in the catchment is 840 mm\

DISADVANTAGES OF EXISTING SYSTEM :

1) Less accuracy
2)low Efficiency

PROPOSED SYSTEM :

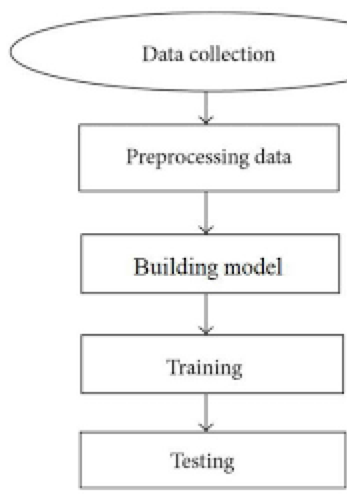
Flood forecasting technique organized method suitably based on available data and an appraisal of rating criteria with an inspired performance. Flood forecasting using real time estimation gives chances of flood value in GUI. Flood estimation using Machine Learning in real time can calculate large data instantly. Comparison between flood modelling by machine learning and stochastic method (i.e. Muskinghum method) gives machine learning is accurate, easy and can be applied for numbers of calculation.

ADVANTAGES OF PROPOSED SYSTEM :

1) High accuracy
2)High efficiency

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE :



V. MODULES

To implement this project we have designed following modules

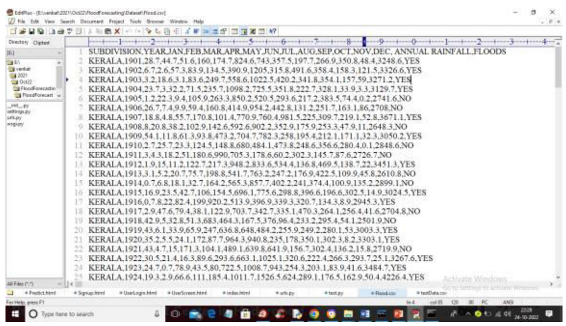
1) New User Signup Here: using this module we will allow user to signup with the application
2) User Login: using this module we will allow user to login to application
3) Preprocess Dataset: using this module we will read flood dataset and then remove missing values and then normalize dataset values and then split dataset into train and test where application use 80% dataset for training and 20% for testing

4) Run Machine Learning Algorithms: using this module we will train all 4 machine learning algorithms such as SVM, Logistic Regression, KNN and MLP and calculate prediction accuracy on test data
5) Forecast Flood: using this module we will upload test data and then MLP will predict flood from that test data.

VI. SCREEN SHOTS

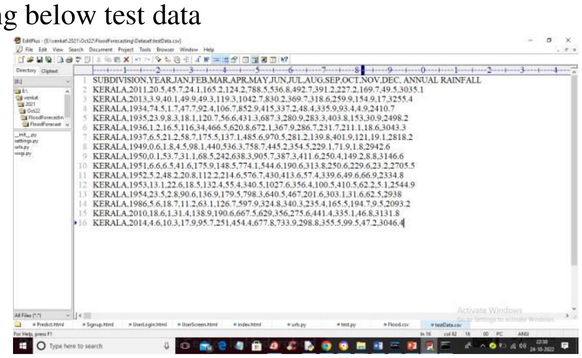
DATASET DESCRIPTION

In this project we are using various machine learning algorithm to predict or forecast flood situation as this is a natural calamity which can cause huge loss of lives and financial assets. Timely and accurate prediction of future floods can help in reduce such loss and to predict flood accurately we have evaluated performance various machine learning algorithms such as SVM, Logistic Regression, MLP and KNN. In all algorithms MLP is giving best performance and to implement this project we have used below flood dataset from KAGGLE website.



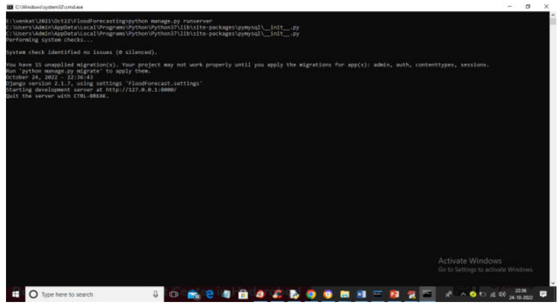
In above dataset first row contains dataset column names and remaining rows contains dataset values. In each row we have monthly and annual rainfall and based on that we have class label as YES (flood occur) and NO (no flood occur). So by using above dataset we will train all algorithms and evaluate their performance in terms of accuracy, precision, recall, FSOCRE, sensitivity and specificity.

To predict flood we are using below test data

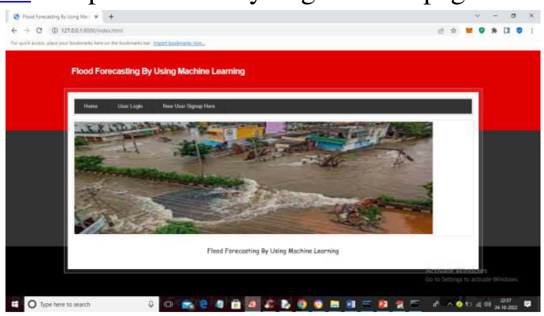


In above test data we have monthly and annually rainfall without flood label and when we apply this dataset on MLP algorithm then it will predict flood will occur or not.

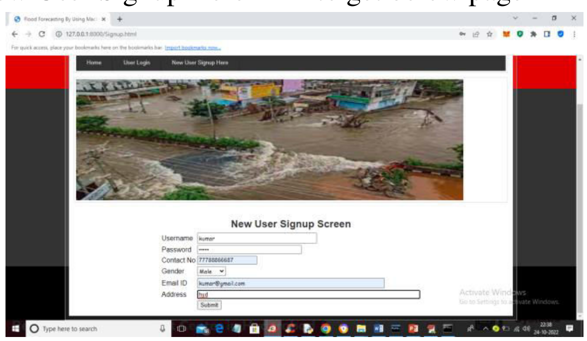
To run project double click on 'run.bat' file to start python DJANGO web server and get below output



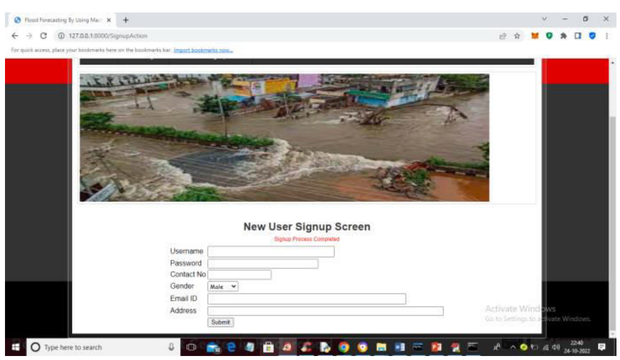
In above screen python DJANGO server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and press enter key to get below page



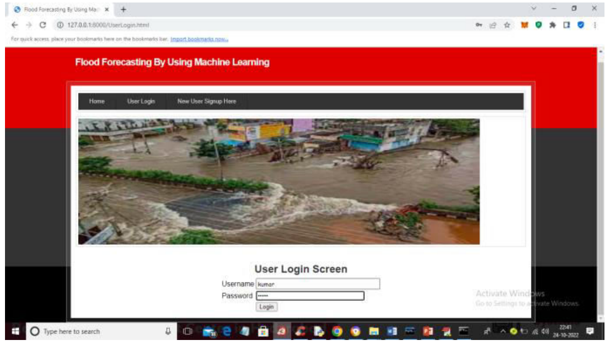
In above screen click on ‘New User Signup Here’ link to get below page



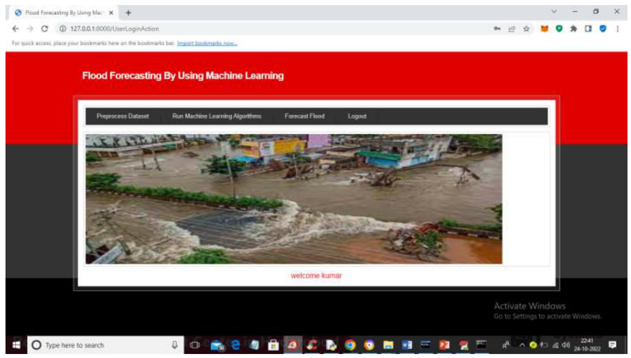
In above screen user is signing up and then click on ‘Submit’ button to complete signup and get below output



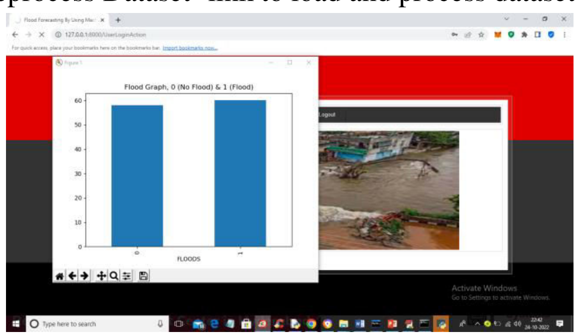
In above screen signup process completed and now click on ‘User Login’ link to get below login screen



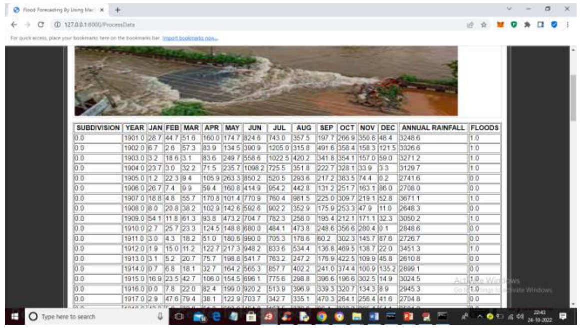
In above screen user is login and after login will get below output



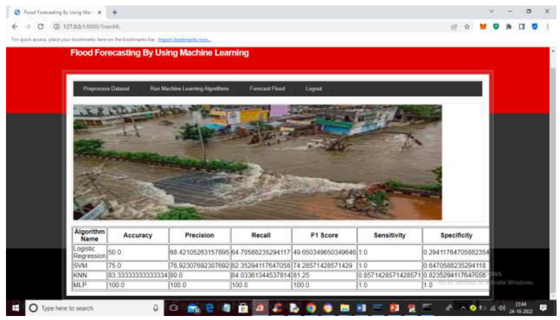
In above screen click on ‘Preprocess Dataset’ link to load and process dataset and get below output



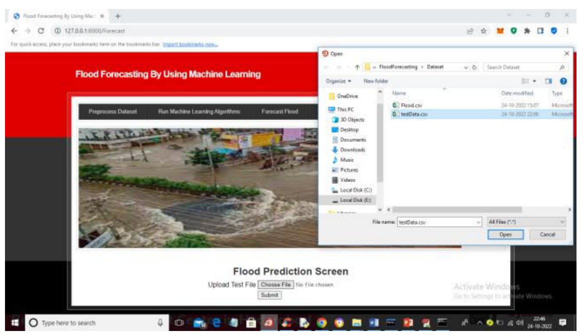
In above screen dataset processing completed and in graph x-axis represents labels as 0 (no flood) and 1 (flood) and y-axis represents number of records in that label and now close above graph to get below output. By using label encoding processing technique we have converted YES and NO to 0 and 1 as machine learning algorithms accept only numeric data



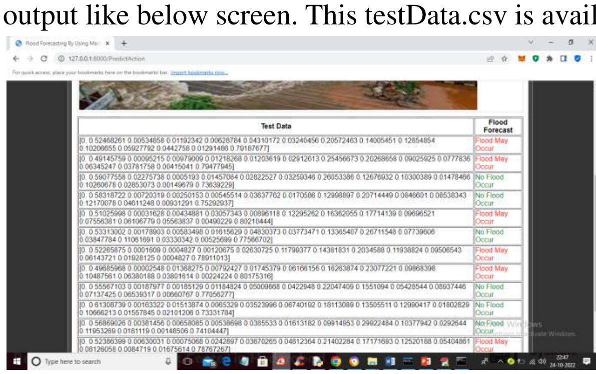
In above screen entire dataset process and loaded and now click on ‘Run Machine Learning Algorithms’ link to train all algorithms and get below output



In above screen in tabular format we can see in all algorithms MLP got highest accuracy as 100% and for each run this accuracy may vary from 95 to 100%. Now algorithms are trained and now click on ‘Forest Flood’ link to get below screen



In above screen select and upload ‘testData.csv’ file and then click on ‘Open’ and ‘Submit’ button to load test data and get prediction output like below screen. This testData.csv is available inside ‘Dataset’ folder



In above screen in first column we can see the Rainfall monthly and annually test data and in last column we can see prediction output as ‘Flood May Occur’ in red colour and ‘No Flood Occur’ in green colour.

VII. CONCLUSION AND FUTURE ENHANCEMENT

The input values at the beginning and end of the jth time period are denoted as Ij and Ij+1, respectively. Similarly, the corresponding outflow values are represented as Qj and Qj+1. Machine learning has the capability to acquire knowledge and enhance a system in a clear and direct way. Machine learning enables computer programs to acquire, process, and analyze data using learning algorithms. Machine learning has the capability to process and analyze massive datasets. An Artificial Intelligent system (AIS) is used to train data in order to enhance the flood forecasting system by providing early stage development warning.

REFERENCES:

1. Pappenberger, F.; Cloke, H.L.; Parker, D.J.; Wetterhall, F.; Richardson, D.S.; Thielen, J. The monetary benefit of early flood warnings in Europe. Environ. Sci. Policy 2015, 51, 278–291.
2. Krzysztofowicz, R. Bayesian system for probabilistic river stage forecasting. J. Hydrol. 2002, 268, 16–40.
3. Todini, E. Role and treatment of uncertainty in real-time flood forecasting. Hydrol. Process. 2004, 18, 2743–2746.
4. Clark, M.P.; Slater, A.G. Probabilistic quantitative precipitation estimation in complex terrain. J. Hydrometeorol. 2006, 7, 3–22.
5. Vrugt, J.A.; Robinson, B.A. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources. 2007, 43.
6. Ebtehaj, M.; Moradkhani, H.; Gupta, H.V. Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap re-sampling. Water Resour. Res. 2010, 46.
7. He, X.; Refsgaard, J.C.; Sonnenborg, T.O.; Vejen, F.; Jensen, K.H. Statistical analysis of the impact of radar rainfall uncertainties on water resources modeling. Water Resources. 2011, 47.
8. Legleiter, C.J.; Kyriakidis, P.C.; McDonald, R.R.; Nelson, J.M. Effects of uncertain topographic input data on two dimensional flow modelling in a gravel-bed river. Water Resour. Res. 2011, 47.
9. Sikorska, A.E.; Scheidegger, A.; Banasik, K.; Rieckermann, J. Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. Hydrol. Earth Syst. Sci. 2012, 16, 1221–1236.
10. Montanari, A.; Koutsoyiannis, D. A blueprint for process-based modeling of uncertain hydrological systems. Water Resour. Res. 2012, 48.
11. Prof.ParthasarathiChoudhary and A. Sankarasubramanian(2009), “River Flood Forecasting Using Complementary MuskingumratingEquations”,Journal of Hydrologic Engineering, Vol. 14, No. 7, July 1, 2009.
12. KoradaHariVenkataDurgaRao, ValaVenkateshwarRao, Vinay Kumar Dadhwal, GandarbhaBehera, and Jaswant Raj Sharma, “A Distributed Model for Real-Time Flood Forecasting In The Godavari Basin Using Space Inputs”, International Journal Disaster Risk Sci. 2011, 2 (3): 31– 40.
13. Sulafa Hag Elsafi(2014) “Artificial Neural Networks (Anns) For Flood Forecasting At Dongola Station In The River Nile, Sudan”, Alexandria Engineering Journal (2014) 53, 655–662.