FLOOD FORECASTING:AN INTELLIGENT SYSTEM USING MACHINE LEARNING

Sandhya ks[1], Ms.Shejina [2]

[1] *Student, Department of Computer Science and Engineering, IES College of Engineering, Kerala, India*
[2] Assistant Professor, *Department of Computer Science and Engineering, IES College of Engineering, Kerala, India*

ABSTRACT

Floods are natural disasters.They are most commonly caused by excessive precipitation and runoffs which lead the adjoining land areas to be submerged by water which causes huge loss to human lives and properties, which includes damaging buildings, bridges ,power supply network and crippling the transportation and bringing economic hardships on the people. In order to decrease the damages caused by the flood, it is essential to implement an intelligent system for the early detection. Over the years, multiple measures have been taken to pre- determine flood warnings which have been implemented using sensor technology and active monitoring of the parameters. This leads to the creation of a wide number of data-sets which can be employed for future purposes.With the availability of data analytics techniques including Machine Learning, the datasets can be directly employed to allow based upon this, and can create a model which helps to predict future outcomes. In this paper,machine learning algorithms like k-nearest neighbor(KNN), decision tree, random forest, and support vector machine (SVM) are used.

Keyword : - *Flood, Machine learning , K Nearest Neighbor, Support vector machine , Decision Tree ,Random Forest*

---

1. INTRODUCTION

Enormous amount of water is called flood. One of the most challenging and difficult problems in hydrology is forecasting of flood. Flood is a situation in which water becomes uncontrollable. Flood may cause due to many reasons. An increase in the risk of a flood hazard can be caused by several factors, including land use changes such as deforestation and rapid urbanization. Demographic pressures also cause the encroachment of informal settlements on hazardous locations in flood plains. Many other factors are likely to be the root causes of flood disasters. Heavy rainfall is a major factor. Flood is one of the most dangerous natural hazards, responsible for loss of lives and damage to properties. A number of regions are subject to monsoons influences and hence face the disaster almost every year.

Forecasting of flood is a difficult procedure. Some of the parameters affecting flood may be amount of rainfall, present river water level, degree of ground saturation, degree of permeable soil etc. In this project, the aim is to prepare appropriate data with correct attributes for the accurate prediction of flood in the regions of India with the help of machine learning techniques.These algorithms are used to train and test the dataset efficiently by using SVM, Random forest, Decision tree and k-nearest neighbor.

2. MOTIVATION

The use of automated learning, artificial intelligence branch is to analyze and predict chance of flood. In many flood prone regions, flood forecasting is one among the few feasible options to manage floods.In the recent years, reliability of forecasts has increased due to the integration of meteorological and hydrological modelling capabilities. It would improve the data collection through satellite observations,and advancements in knowledge and algorithms for analysis .It can depend subjective data from informative survey.Early flood warning systems using sensor systems, Geographical Information Systems (GIS) etc couldn't identify chances of flood properly. So our developed system can accurately predict the chances of flood more efficiently.

3. LITERATURE SUEVEY

Flood Detection Using Gradient Boost Machine Learning Approach proposed by A Yovan Felix and T.Sasipraba [1] presented Flood Detection mechanism using the Gradient Boost Algorithm which will be used to classify the data sets and perform regression on it to produce the best outcomes from the datasets we will use to train it, to create a weak prediction mode brought the concept based on a Decision Tree. The outcome can henceforth be used to display it to the concerned authorities who can employ preemptive actions to tackle the threat. This approach is developed to be better suited in such ends providing predictions with high accuracy and additionally employs various other technologies like Remote Sensing and Sensor Technology to develop accurate datasets required to train the model. The Flood Detection mechanism follows the procedure of determining the level of surface water, rainfall characteristics and runoff to determine the chances of flood and can be proposed to local meteorological and disaster management agencies as a novel architecture to implement an intelligent machine to predict the chances of a flood in case of an inevitability.

Automated Change Detection in Satellite Images Using Machine Learning Algorithms for Delhi, India presented by Abhishek Bhatt,S K Ghosh and Anil Kumar[2] analyzed effectiveness of the three types of unsupervised Machine learning algorithms for change detection to detect the change in some of the dominant classes in an urban area, such as, vegetation, water bodies. Three indices namely Normalized Difference Built up Index (NDBI), Modified Normalized Difference Water Index (MNDWI) and Modified Soil Adjusted Vegetation Index (MSAVI2) have been generated from the Landsat data. With the use of indices extracted a large reduction in data dimensionality is possible. Three algorithms, namely, K-Means, FCM(Fuzzy-c-means) and EM(Expectation Maximization) have been used. K-means clustering is a widely used ML algorithm mainly used to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Fuzzy C-Means Clustering is one of the most widely used methods.It is for unsupervised classification in remote sensing discipline. The algorithm is an iterative clustering method that produces an optimal number of partition by minimizing the weighted within group sum of squared error objective function. The EM algorithm aims to find the maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. The EM algorithm also used to find the maximum likelihood parameters of a statistical model. These models consist of latent variables in addition to unknown parameters and known data observations. In fact, EM offers an iterative approach of fill-in the missing values by single imputation.

Flood Forecasting System Based on Integrated Big and Crowd source Data by Using Machine Learning Techniques presented by Supattra Puttinaovarat and Paramate Horkaew [3]explained a novel flood forecasting system based on fusing meteorological ,hydrological, geospatial, and croud source big data in an adaptive machine learning framework .Firstly, meteorological and hydrological data were acquired from the Global Flood server, called GLOFAS. Meteorological data consisted of accumulated precipitation, and the probability of precipitation at different levels, pre- dicted daily. The second source of data is rainfall forecasting, which was acquired from big data repository, managed by TMD(Thailand Meteorological Department). Thirdly, area-specific geospatial data consisted of height above sea level, slope, land use and land cover, 10-year repeating flood and flow direction. The fourth group was crowd sourced data gathered from the public. They were further divided into two parts. Data in the first part were used in training the forecasting system. They were essentially real-time data accounting actual incidents. These real-time data consisted of area specific rainfall intensity levels, continuing rainfall durations, and drainage ability. The second part is verification data, characterizing floods happening in given areas.One of the contribution is employing croud source data ,not only in training but also in verification.Forecasting is performed based on these data learned by modern ML strategies. They were decision tree, RF, NaïveBayes, MLP and RBF and ANN ,SVM and fuzzy logic.

Flood Eye: Real-time Flash Flood Prediction System for Urban Complex Water Flow presented by Kei Hiroi and Nobuo Kawaguchi[4] explained a water-level sensor system that has multiple functions: real-time river-level monitoring and high accurate flood prediction for urban complex water flow (CWF) flooding caused by Localized Heavy Rain.A detection scheme for CWF by developing a water-level sensor system that works in various installation environments using infrared image processing with both low installation and operation cost.A CWF prediction system that produces accurate and early predictions using a Linear Regression (LR) of deep learning approach with data assimilation. The effect of knowing when river water rises is an important factor in avoiding vulnerable areas in situations of increasing Localized Heavy Rain (LHR) caused by global climate change.To predict higher accurate water level rising, we use rainfall values to adapt CWF situations by calculating two patterns of rainfall value: 10- minute and total amounts.Considering the rainfall patterns of LHR is for precision with respect to other prediction methods. The implementation of the Flood Eye monitoring function and an examination of its effectiveness are described here.Sensor data that were collected from sites where LHR events previously occurred in Fukui, Japan, were used for evaluation. Flood Eye was installed at 13 locations, collected and calculated water level and riverbank height in 1-minute intervals over a 30-minute duration on an LHR day.This sensor system with multiple functions: water level monitoring and flash flood prediction that produces accurate and early predictions.The developed a low-cost water-level sensor system with auto-configuration and a stable monitor/prediction scheme in various urban riverside environments for the purpose of high- resolution CWF prediction.

Short-term Water Level Prediction using Different Artificial Intelligent Models proposed by Lizhen Lu,Shuyu Zhang, Jianjun Yu and Honjie Zhou[5] and they explained artificial neural networks (ANN), support vector machine (SVM) and adaptive neuro fuzzy inference system (ANFIS) are selected for comparison. The 1- to 4-hour ahead forecasting based on previous 2-hour inputs.ANFIS showed better ability of avoiding information noise with different lags of inputs. SVM is relatively more robust during peak values under extreme typhoon events.Overall, this study shed some lights on the best practices in using various artificial intelligent methods for prediction of hourly water level with different model settings. ANN consists of a large number of parallel processing neuros,working independently and connecting to each other by weighted links.The feed forward neural network (FFNN), with one input layer, one or more hidden layer and one output layer, is included in this study. SVM is used to find an optimal separating hyperplane to divide the samples into two groups through maximizing the distance between them.Adaptive Neuro Fuzzy Inference System (ANFIS) combines the advantages of ANN and FIS. It expresses the algorithm in fuzzy language and can adjust the parameters of fuzzy rules and membership functions auto.The results showed that ANN using samples in heavy rainfall period performed better than that in using all data values.4-hour ahead forecasting based on previous 2-hour inputs is the suggested modeling scenario for ANN, SVM and ANFIS.matically during training and learning process.

Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework proposed by Jeo Marlou A. Opella and Alexpander A Hermandez [6]developed a Fusing ConvNet, a feedforward neural networks that specialize in image processing and prediction with SVM, a supervised machine learning for classification and regression analysis.This system aims to exploit the data available from the Geographical Information System (GIS) and the technology advancement in the modern world in producing a reliable flood susceptibility and probability map.In this method is distributed into three main stages: (i) extraction: In this stage, an image will be processed using dilated convolution and deconvolution network, which will yield different mapping probability and prediction. All maps that contain identical resolution from the initially inputted images will be concatenated generating a feature vector which can signify the input image. (ii) Learning: In this stage, the SVM accepts the featured vector above, as well as the flooding ground-truth maps for all training data. Then, it separately processes every pixel of these images, learning which and when each classifier is better; and (iii) Prediction: this is final stage, accepts the testing images of the feature vectors using the trained SVM, and outputs the enriched prediction map for each test image. The main model which we used for training and inference purpose is described here. A deep neural network has the potential to aid surficial geology mapping where experts can easily modify and speed up the development process. Previous studies have confirmed that combining CNN, a deep feedforward artificial neural network generally applied to computer vision problem such as image processing and SVM which used binary classification to obtain optimal hyperplane got very high accuracy.

Early Flood Risk Assessment using Machine Learning: A Comparative study of SVM, Q-SVM, K- NN and LDA presented by Talha Ahmed Khan, Zeeshan Shahid, Muhammad Alam, M.M Su'ud and Kushsairy Kadir [7]explained A dynamic system for the identification of run offs involves the computation of water peak, rainfall velocity, Global Positioning System-Precipitable Water Vapor (GPS PWV), wind speed, orientation, complex levels of river, land humidity, oceanic basement pressure and flash flood color with authentic cognizance algorithms.In this, classification approaches like Linear Support vector machine, Quadratic Support vector machine, K-nearest neighbor and Linear discriminant analysis have been implemented to classify the true positive event of flash floods accurately and precisely. Flood risk prediction system have been developed using machine learning that eliminates the extra efforts involved in manual estimation and provides the desired precision for decision-making.The system has access to flood information and the weather information, therefore it calculates the event before time that can help concerned authorities plan.

Automatic Detection of Flood Severity Level from Flood Videos using Deep Learning Models proposed by Kanishk Lohumi and Sudip Roy[8] presented the video of a flooding event as input and determines its severity level.In this, create a dataset of flood videos due to the unavailability of such special kind of dataset.Then evaluated on this dataset and compared with a baseline convolutional neural network (CNN) based model. Categorized the video clips into four categories as low, moderate, high, very high, which signifies the severity of the flood. A general idea is to use CNN architecture to get the frame level representation of the videos and then use them to infer some information about the video by averaging the frame level features. The main model which we used for training and inference purpose is described here. We used a GRU as a basic unit which is unrolled number of times equal to the number of frames present in the video clip on which training or inference is performed. The GRU unit captures the long or short term dependencies present in the sequential data at different time-steps.

Performance Evaluation of Different Machine Learning Based Algorithms for Flood Prediction and Model for Real Time Flood Prediction[9]proposed by Chinmayiee kinage,Abhishek Kalgutkar, Amruta Parab, Sejal Mandora and Sunita Sahu developed to determine which works best among the machine learning algorithms and the influential parameters.In this flood prediction model for Mumbai using machine learning and developed an Android application for the same.. Support Vector Machine, Naive Bayes, Logistic Regression, Random forest and Recurrent Neural Networks are used. As the collected dataset is huge, machine learning helps to handle multidimensional data in dynamic environment. The android application is integrated with a RNN model in the backend. The factors considered for prediction are precipitation, drainage, humidity, temperature, pressure and wind speed. The application also has the functionality of sending an alert to other citizens via mail, messages, etc.More reliable results got from RNN.

Real-time WSN Based Early Flood Detection and Control Monitoring System[10] proposed by Tibin Mathew Thekki and Dr.N.Prabakaran developed a real time Wireless sensor network based early flood detection and control monitoring system designed with a function of real time monitoring ,guaranteeing connectivity in low cost. This system collects data as images from CMOS image sensors through wireless sensor nodes which transmit these images to remote monitoring centre via Zigbee network and GSM network. Zigbee router collects and transmits the images captured by the sensor to gateways. GSM network use internet to transfer these data to remote monitoring centre. In this design early flood detection and control monitoring system studied, which should record and give proper warnings in real time.This used image comparing algorithms SIFT and BRIEF. Scale Invariant Feature Transform (SIFT) is a feature detector. SIFT has proven to be very efficient in object recognition applications. Binary Robust Independent Elementary Features (BRIEF) is another algorithm requires less complexity with almost similar matching performance.

Automated Wireless Flood Warning System in Remote Areas[11] proposed by C.Srihari, S.Prashanthi, V.Sriranjani and Sobitha Ahila.S developed e a safety device, which measures the water level in the water body and alerts people if the level crosses the set threshold level, is being developed. A sensor is placed in the water body at the threshold level. The wireless water level sensor is placed in the measurement scale that is used to measure the water level. The measured water levels are continuously sent to the raspberry pi with the help of Radio Frequency signals through the XRF Module Transceiver. XRF is an extremely popular, easy to use, RF serial data module. In the monitoring phase,in order to establish a connection through sensors, we need RF Transmitters Antenna, Water Sensors, Battery and Cables.In the processing phase, The radio frequency signals from the water level sensor are received by the XRF module. Omnidirectional antennas are used here. The main motto of this work is to save everyone by sending vital information regarding a disaster like flood well before so that people will be ready before the actual calamity.

An Improved Flood Warning System using WSN and Artificial Neural Network[12]proposed by Prof.(Dr.) J.K.Roy, Dola Gupta and Sanjay Goswami developed Flood Warning System. The system consists of M number of wireless sensor nodes (WSNs) deployed at several locations along the river bed of Damodar, India. Six nodes taken together constitute a cluster. The six parameters considered are Water velocity (m/s), Water level (m), Air moisture or Humidity (%), Wind velocity (km/h), Rainfall (mm) and Atmospheric pressure (atm). These features were divided equally into 4 classes based on the prioritized combination of the feature values. The input vectors after simulation produces one of the four output vectors as Alarming, Low Flood, High Flood and Very High Flood. The model considered for ANN simulation is designed with 6 input vectors/features, 20 neurons in the hidden layer (layer 1) and 4 output vectors in layer 2. The six input parameters given to the neural network is the six hydrological parameters measured by each WSN in a cluster.

Application HEC-RAS Model to Simulate the Flood Wave Due to Dam Failure[13] proposed by Layla A. Mohammed, Saleh I. Khassaf and Kareem R. Al-Murshidi developed the numerical software (HEC-RAS) which is based on Saint-Venant equations was applied to get the main parameters of the flood wave at certain cross sections. The river schematics are essential for any set of geometric data to create the HEC-RAS model, this sketch must be created before any other data is performed.Contraction and Expansion Coefficients calculated for the change in the shape of the river area is one of the main reasons for the occurrence of energy losses. Manning Coefficient (n)is a parameter depends on several factors, such as the bed roughness, vegetation, irregularities, and meandering of the stream.Statistical analysis is performed to develop the best formulas for estimating Max. elevation and Max. discharge depending on distance and time. Also, it is found that the flood wave takes long time to achieve the highest level for stations located farther from the dam site, thus these stations were considered an accumulation zone of the flood.

The proposed system of flood prediction detection system consist of the following modules.

3.1 Data Preparation

The aim is to prepare and assemble the rainfall data in various region of India. Mainly rainfall in Kerala. After a thorough study of the flood-prone area, the dataset is prepared and labeled accordingly.

3.2 Data Pre-processing

This is the most pivotal step concerning prediction as inconsistent data can give rise to inaccurate results. The data can be obtained from various resources available in big data. The important steps are:

Import Libraries Numpy and Pandas where the libraries imported for processing the dataset. Numpy is the fundamental package for scientific computing with python. Pandas is an open- source high-performance library for data manipulation and analysis.Import dataset is one of the important step.The prepared data is compiled into a .csv(Comma Separated Values) file and imported using pandas.There will have a chance of missing values .This will lead inaccurate inferences regarding the data. Removal of data with missing values is not an efficient method. Instead of missing values, we are putting null values . There are several factors to consider in the data cleaning process. Elimination of irrelevant data columns, Bad labeling of data, same category occurring multiple times. Data cleaning process need carefully consideration because the influence the results.

3.3Model selection and prediction

Model selection or algorithm selection phase is the most exciting and the heart of machine learning. It is the phase where select the model which performs best for the data set at hand. Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset. Model selection is a process that can be applied both across different types of models .Different kinds of machine learning algorithms are used to predict flood. They are SVM(support vector machine),Random forest , KNN and Decision tree etc.

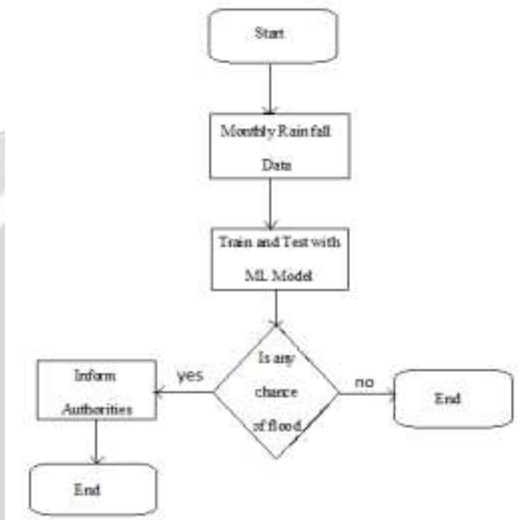

Fig -1: Flow chart of flood prediction system

4. CONCLUSIONS

Flood prediction is extremely complex and challenging process. But flood forecasting system helps to predetermine flood events. By developing these systems, meteorological authorities and disaster management teams can control incoming floods, and the evacuation of the affected people to the safer places. The chosen methodology is to evaluate the performance of system in order to determine the most appropriate predictive model. The developed system mainly concentrated the Rainfall data of different regions in India mainly Kerala. This study compares the accuracy score of decision tree, KNN, SVM, and Random Forest for flood prediction. The results of this study indicates that the Random forest which is the most effective machine learning algorithm with accuracy score 83%.

5. ACKNOWLEDGEMENT

I would like to convey my heartfelt gratitude towards my guide Ms.Shejina for her constant guidance,encouraging help and inspiring words.Iam thankful to the department of computer science engineering for their support.

6. REFERENCES

[1]. D Han, L.Chan and N Zhu:" Flood forecasting using support vector machines", in 2007 in Journal of Hydroinformatics.
[2 Amir Gosavi, Pinar Ozturk,Kwok-wing Chau: " Flood prediction using machine learning models, 27 October 2018, MDPI Journal
[3]. A Yohan Felix, T. Sasiprabha: "Flood detection using gradient boost machine learning approach", in December 11-12 2019, International Conference on Computational Intelligence and Knowledge

[4]. P.C Nayak, K.P Sudheer, D.M Rangan, K.S Ramasastri: " Short term flood forecasting with a neurofuzzy model, in 5 April 2005, Water Resources Research, Volume-41.
[5]. A Yohan Felix, T. Sasiprabha: Flood detection using gradient boost machine learning approach, in December 11-12 2019, International Conference on Computational Intelligence and Knowledge Economy.
[6]. Tibin Matthew Thekkil, Dr. N. Prabhakaran: "Real time wsn based flood detection and control monitoring system", 2017, International Conference on Intelligent Computing, Instrumentation and Control Technologies.
[7]. Abhishek Bhatt, S.K Ghosh, Anil Kumar: "Automated change detection in satellite images using machine learning algorithms for delhi, india"
[8]. Supatra Puttinaoporavat, Paramate Horkaew: Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques, 10.1109/ \PRP6105021450
[9]. Joe Malau A.Opella, Alexander Hernandez: "Developing a flood risk assessment using support vector machine and convolutional neural network: a conceptual framework", in March 9, 2019, 15 th International Colloquium on signal processing and Applications
[10]. C. Srihari, S. Prashanthi, V. Sriranjani, S. Ahila.S. "Automated wireless flood warning system in remote areas", Proceedings - 2016 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development, TIAR 2016, pp. 181- 187
[11]. J.K. Roy, D. Gupta, S. Goswami, "An improved flood warning system using WSN and artificial neural network", 2012 Annual IEEE India Conference, INDICON 2012, pp. 770-774
[12]. Kanishk Lohumi, Sudip Roy: "automatic detection of flood severity level from flood videos using deep learning models", 2018, 5th International Conference on Information and Communication Technologies For Disaster Management
[13]. Talha Ahmed Khan, Zeeshan Shahid, Muhammad Alam, M.M Su'ud Kushsairy Kadir: "Early flood risk assessment using machine learning: a comparative study of svm, q- svm, knn and lda", 2019, 13 th International Conference on Mathematics, Actuarial Science, Computer Science And Statistics
[14]. Kei Hiroi, Nobuo Kawaguchi:"Floodeye: real-time flash flood prediction system for urban complex water flow", 2016 IEEE Sensors Conference
[15]. Djati Mardiatno, Nurul Khakim, Tri Kuntoro Priyambodo: "Identification of flood prone area using remotely sensed data", 2015, IEEE International Conference on Aerospace Electronics and Remote Sensing Technology