International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Machine Learning with Flood Prediction.

Yogesh Arjun Bodke[1], Sachin Haribhau Jagtap[2], Neha Sanket Supekar[3], Vaishnavi Rajesh Suvernkar[4]
[1,2,3,4,] *Civil Engineering B.TECH., G H Raisoni College of Engineering and Management,*
Wagholi, Pune, Maharashtra, India

Abstract -: Floods are the most common natural disasters that are highly complex to model. They can cause destruction and devastation of natural life, agriculture, infrastructure and properties, every year. There are many researches on the advancement of flood prediction models. These models have contributed well to risk reduction, minimisation of the loss of human life and reduction of property damage associated with flood. Here, we use Machine Learning models for prediction. ML methods provide better performance and costeffective solutions. Some of the ML algorithms which were reported as effective for both short-term and longterm flood forecasts are Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Support Vector Regression (SVR). Here, in this report we introduce the most promising prediction method for both short-term and long-term floods.
Flood prediction is one of the most challenging and difficult problems in hydrology. Flood disaster had great impact on city development like reduction in economic condition and life loses. The aim of this is to discovering more accurate and efficient prediction model. The main contribution of this is to demonstrate the state of the art of ML models in flood prediction and to give insight into the most suitable models. Early Flood Warning System (EFWS) are promising curative against flood hazards and losses. Machine learning is the centerpiece for building a satisfactory early flood warning system.
The aim of this project is to discover more accurate and efficient prediction models. The main contribution of this is to demonstrate the state of the art of ML models in flood prediction and to give insight into the most suitable models. other relevant flood prediction variables. In the process of

***Key Words*:** Flood forecasting, Machine learning, hydrological research aspect. Prediction using machine-Support Vector Machine, Rainfall runoff model, Stream low. learning algorithms is effective due to its ability to utilize from various sources and classify and regress it into flood.

1. INTRODUCTION

Flood Prediction is one of the most challenging and difficult problems in hydrology. Flood disaster had great impact on city development like reduction in economic condition and life losses. In many regions the flood forecasting technique is the reliable technique for more accurate flood prediction. The flood formation system using ML algorithm is a convenient tool for managing floods in the integrating analytics and also machine learning algorithms for more dynamic flood risk visualization and prediction.
The aim is discovering more accurate and efficient prediction models. The main contribution of this is to demonstrate the state of the art of ML models in flood prediction and to give insight into the most suitable models.
Early Flood Warning System (EFWS) are promising curative against flood hazards and losses. Machine learning is the centerpiece for building a satisfactory early flood warning system. This summarizes the machine learning methods proposed in these special issues for flood forecasts and prediction with their proper advantages.
The changing patterns and behaviors of river water levels that may lead to flooding are an interesting and practical research area. They are configured to mitigate economic and societal implications brought about by floods. Support Vector Machine (SVM) are machine learning algorithms suitable for predicting changes in levels of river water, thus detection of flooding possibilities. The two algorithms employ similar hydrological and flood resource variables such as precipitation amount, river inflow, peak gust, seasonal flow, flood frequency, and

Flood prediction is needed for developing appropriate measures to control flood risk, mitigate flood hazard, evacuate people from flood hazard areas, determine insurance premiums, and manage environmental and water resources systems. Data-driven techniques/models have gained a significant attention for flood forecasting in recent years, of which SVM is greatly popular in flood modelling; it is a supervised learning machine which works based on statistical learning theory and the structural risk minimization rule. SVM is used to predict a quantity forward in time based on training from past data emerged as alternative ML methods to ANNs, with high popularity among hydrologists for flood prediction. Thus, they are applied in numerous flood prediction cases with promising results, excellent generalization ability, and better performance, compared to ANNs, e.g., extreme rainfall, precipitation, rainfall-runoff, reservoir inflow, streamflow, flood quantiles, flood time series, and soil moisture. SVM is considered as the most widely applied state-of-the-art machine learning method. It is mainly used for classification problem. The mechanism of SVM is applying the principle of margin calculation. It basically evaluates margins between the classes. The margins are evaluated in such a way that the gap between the margin and the classes is maximum and thus, classification error is minimized.

The frame work includes –
1] Data Collection,
2] Data transmission,
3] Data Analysis & forecast formulation, 4] Dissemination (i.e., Publicize).

Data collection is defined as the procedure of collecting, measuring and analysing accurate insights for research using standard validated techniques**.** A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research.

Data transmission can be analogue and digital but is mainly reserved for sending and receiving digital data. It works when a device or piece of equipment, such as a computer, intends to send a data object or file to one or multiple recipient devices, like a computer or server. The digital data originates from the source device in the form of discrete signals or digital bit streams. These data streams/signals are placed over a communication medium, such as physical copper wires, wireless carriers and optical fibre, for delivery to the destination/recipient device. Moreover, each outward signal can be baseband or passband.

Analysis refers to breaking a whole into its separate components for individual examination. Data analysis is a process for obtaining raw data and converting it into information useful for decision-making by users. Data is collected and analysed to answer questions, test hypotheses or disprove theories. Also, it is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

Dissemination is a planned process that involves consideration of target audiences and the settings in which research findings are to be received and, where appropriate, communicating and interacting with wider policy and health service audiences in ways that will facilitate research uptake in decision-making processes and practice.

1.1 MACHINE LEARNING

Machine learning is defined as the field of study that gives computers the ability to learn without being explicitly programmed. Machine learning (ML) is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the extract information from the data. In that case, we apply machine learning. With the abundance of datasets available, the demand for machine learning is in rise. Many industries apply machine learning to extract relevant data. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves without being explicitly programmed. Many mathematicians and programmers apply several approaches to find the solution of this problem which are having huge data sets. Natural disasters, such as landslides, serious floods, fires, volcanic eruptions and the damage they cause, are global problems, which incur a heavy cost in terms of human lives and financial losses. Every year people are affected by the problem of flooding, and it is expected that this number will increase in the years ahead. The regions that suffer most from floods are developing countries and urban areas. Where the local climate has changed in the last few years, resulting in serious flooding.

1.2 LITERATURE SURVEY

Floods are among the most destructive natural disasters, which are highly complex to model. The research on the advancement of flood prediction models contributed to risk reduction, policy suggestion, minimization of the loss of human life, and reduction of the property damage associated with floods. To mimic the complex mathematical expressions of physical processes of floods, during the past two decades, machine learning (ML) methods contributed

highly in the advancement of prediction systems providing better performance and cost-effective solutions. Due to the vast benefits and potential of ML, its popularity dramatically increased among hydrologists.

The literature where ML models were benchmarked through a qualitative analysis of robustness, accuracy, effectiveness, and speed are particularly investigated to provide an extensive overview on the various ML algorithms used in the field. The performance comparison of ML models presents an in-depth understanding of the different techniques within the framework of a comprehensive evaluation and discussion. As a result, this paper introduces the most promising prediction methods for both long-term and short-term floods.

Furthermore, the major trends in improving the quality of the flood prediction models are investigated. Among them, hybridization, data decomposition, algorithm ensemble, and model optimization are reported as the most effective strategies for the improvement of ML methods. This survey can be used as a guideline for hydrologists as well as climate scientists in choosing the proper ML method according to the prediction task. The studies that show easiest way to minimize the error in Nam model, in practical use HD model to analyze. An essential countermeasure towards better flooding management is to forecast flood water levels in the realtime manner. Most existing early warning systems (EWS) in Thailand contain a lot of miscalculations when they face with real situations. Towards prediction improvement, this presents hydrological modeling augmented with alternative five machine learning techniques; linear regression, neural network regression, Bayesian linear regression and boosted decision tree regression. As the testbed system, the so-called MIKE-11 hydrologic forecasting model, developed by Danish Hydraulic Institute (DHI), Denmark, is used. To test error reduction in runoff forecasting, the water-level records during 2012-2016 data are used for training and the derived model is tested on the record of 2017, in the experiments.

Natural flood in monsoon and seasonal deluge almost in every year. The study of the area in south Thailand. The forecasting was performed strategies based on various modern machine learning technique. They include DT, ANNs, SVM, SVR and Fuzzy logic. The evaluation result on studied areas indicated that the system could forecasted flood event highly accurately. Bangladesh is a flood prone country and extreme floods inundate more than half of the country's landmass almost every year. So, initially co-axial correlations, gauge-togauge statistical co-relation relationships, and Muskingum-Conge routing method were used for forecasting of water levels in advance.

2.METHODOLOGY

I. Support Vector Machine [SVM] SVM is greatly popular in flood modelling; it is a supervised learning machine which works based on statistical learning theory and the structural risk minimization rule. The training algorithm of SVM builds models that assign new no probabilistic binary linear classifiers, which minimize the empirical classification error and maximize the geometric margin via inverse problem-solving. SVM is used to predict a quantity forward in time based on training from past data. SVMs are today known as robust and efficient ML algorithms for flood prediction. SVM and SVR emerged as alternative ML methods to ANNs, with high popularity among hydrologists for flood prediction. Thus, they are applied in numerous flood prediction cases with promising results, excellent generalization ability, and better performance, compared to ANNs, e.g., extreme rainfall, precipitation, rainfall-runoff, reservoir inflow, stream flow, flood quartiles, flood time series, and soil moisture.

II. Decision Tree [DT]

DT is one of the contributors in predictive modelling with a wide application in flood simulation. DT uses a tree of decisions from branches to the target values of leaves. In classification trees (CT), the final variables in a DT contain a discrete set of values where leaves represent class labels and branches represent conjunctions of features labels. When the target variable in a DT has continuous values and an ensemble of trees is involved, it is called a regression tree (RT). Regression and classification trees share some similarities and differences. As DTs are classified as fast algorithms, they became very popular in ensemble forms to model and predict floods. The random forests (RF) method is another popular DT method for flood prediction. RF includes several tree predictors. Each tree creates a set of response predictor values associated with a set of independent values. Furthermore, an ensemble of these trees selects the best choice of classes.

III. K-Nearest Neighbour [KNN]

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique-NN algorithm assumes the similarity between the new

case/data and available cases and put the new case into the category that is most similar to the available categories-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems'-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

IV. LINEAR REGRESSION [LR]

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis. Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are nonlinearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine. Linear regression has many practical uses. Most applications fall into one of the following two broad categories: If the goal is prediction, forecasting, or error reduction, [clarification needed] linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.

3. METHOD AND OUTLINE

This survey identifies the state of the art of ML methods for flood prediction where peer-reviewed articles in top-level subject fields are reviewed. Among the articles identified, through search queries using the search strategy, those including the performance evaluation and comparison of ML methods were given priority to be included in the review to identify the ML methods that perform better in particular applications. Furthermore, to choose an article, four types of quality measure for each article were considered, i.e., source normalized impact per paper (SNIP), Cite Score, SC Imago journal rank (SJR), and h-index. The papers were reviewed in terms of flood resource variables, ML methods, prediction type, and the obtained results. The applications in flood prediction can be classified according to flood resource variables, i.e., water level, river flood, soil moisture, rainfall– discharge, precipitation, river inflow, peak flow, river flow, rainfall–runoff, flash flood, rainfall, streamflow, seasonal stream flow, flood peak discharge, urban flood, plain flood, groundwater level, rainfall stage, flood frequency analysis, flood quantiles, surge level, extreme flow, storm surge, typhoon rainfall, and daily flows. Among these key influencing flood resource variables, rainfall and the spatial examination of the hydrologic cycle had the most remarkable role in runoff and flood modeling. This is the reason why quantitative rainfall prediction, including avalanches, slush flow, and melting snow, is traditionally used for flood prediction, especially in the prediction of flash floods or short-term flood prediction. However, rainfall prediction was shown to be inadequate for accurate flood prediction. For instance, the prediction of streamflow in a long-term flood prediction scenario depends on soil moisture estimates in a catchment, in addition to rainfall. Although, high-resolution precipitation forecasting is essential, other flood resource variables were considered. Thus, the methodology of this literature review aims to include the most effective flood resource variables in the search queries. A combination of these flood resource variables and ML methods was used to implement the complete list of search queries. Note that the ML methods for flood prediction may vary significantly according to the application, dataset, and prediction type. For instance, ML methods used for short-term water level prediction are significantly different from those used for long-term streamflow prediction.

4. OBJECTIVE

To calculate flood forecasting by using the Machine Learning technique.
To determine a prediction model by using ML algorithm of Kerala region.
To identify the causes of flood

5. RESULT

The expected results are accurate flood prediction by using the available rainfall data. When inputs such as location, rainfall, climate are given it shows the chance of occurrence of flood. And also analyses the more accurate method of flood prediction and that detected method is linear regression

6. CONCLUSIONS

Flood prediction helps to inform the public about the possibility of an incoming flood disaster. Accurate prediction of flood is important to avoid or reduce the impact caused by the disaster. This paper proposes a machine learning algorithm to calculate the occurrence of flood. In order to effectively obtain the required output, the project is done by using the software 'Jupyter Notebook'.

6. REFERENCES

[1] Chang, F., Hsu, K., & Chang, L. (2019). Flood Forecasting Using Machine Learning Methods. In *Flood Forecasting Using Machine Learning Methods*. https://doi.org/10.3390/books978-3-03897-549-6

[2] Chang, F., Hsu, K., & Chang, L. (2019). Flood Forecasting Using Machine Learning Methods. In *Flood Forecasting Using Machine Learning Methods*.
http://www.mdpi.com/books/pdfview/book/1151

Flood Prediction Using Machine Learning Models: Literature Review 10.3390/w10111536 Mosaic, A., Oz Turk, P., & Chua, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water (Switzerland)*, *10*(11), 1–40.
https://doi.org/10.3390/w10111536

[3] Rao, B. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174–1179. https://doi.org/10.21275/ART20203995

[4] P. Goswami and Srividya, "A novel Neural Network design for long range prediction of rainfall pattern," Current Sci. (Bangalore), vol. 70, no. 6, pp. 447-457, 1996.

[5] C. Venkatesan, S. D. Raskar , S. S. Tambe , B. D. Kulkarni , and R. N. Keshavamurty , "Prediction of all India summer monsoon rainfall using Error-Back-Propagation Neural Networks," Meteorology and Atmospheric Physics, pp.
225-240, 1997.

[6] A. K. Sahai, M. K. Soman, and V. Satyan, "All India summer monsoon rainfall prediction using an Artificial Neural Network," Climate dynamics, vol. 16, no. 4, pp. 291-302, 2000.

[7] N. S. Philip and K. B. Joseph, "On the predictability of rainfall in Kerala-An application of ABF neural network," Computational ScienceICCS, Springer Berlin Heidelberg, pp. 1-12, 2001.

[8] N. S. Philip and K. B. Joseph, "A Neural Network tool for analysing trends in rainfall," Compute. & Geosci. vol. 29, no. 2, pp. 215-223, 2003.

[9] S. Chattopadhyay and G. Chattopadhyay, "Comparative study among different neural net learning algorithms applied to rainfall time series, "Meteorological applicant.
vol. 15, no. 2, pp. 273-280, 2008.

[10] P.Guhathakurta, "Long lead monsoon rainfall prediction using deterministic Artificial Neural Network model,"
Meteorology and Atmospheric Physics 101, pp. 93-108, 2008.

[11] . L. Wu, K. W. Chau, and C. Fan, "Prediction of rainfall time series using Modular Artificial Neural Networks coupled with data pre-processing techniques," J. of hydrology, vol. 389, no. 1, pp. 146- 167, 2010. [13] K. K. Htike and O. O. Khalifa, "Rainfall forecasting models using Focused Time-Delay Neural Networks," Compute. And Commun.
Eng. (ICCCE), Int. Conf. on IEEE, 2010.

[12] Danso-Amoako, E.; Scholz, M.; Kalimeris, N.; Yang, Q.; Shao, J. Predicting dam failure risk for sustainable flood retention basins: A generic case study for the wider greater manchester area. Comput. Environ. Urban Syst.
2012, 36, 423–433.

[13] Xie, K.; Ozbay, K.; Zhu, Y.; Yang, H. Evacuation zone modeling under climate change: A data-driven method. J. Infrastruct. Syst. 2017, 23, 04017013.

[14] Pitt, M. Learning Lessons from the 2007 Floods; Cabinet Office: London, UK, 2008.

[15] Lohani, A.K.; Goel, N.; Bhatia, K. Improving real time flood forecasting using fuzzy inference system. J. Hydrol. 2014, 509, 25–41.

[16] Saiesh Naik, Srushti Ashok Patil, Anang Verma, Anil Hingmire, 2021, Flood Prediction using Logistic Regression for Kerala State, INTERNATIONAL JOURNAL
OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)
NTASU – 2020 (Volume 09 – Issue 03).

[17] Mosavi, Amir, Pinar Ozturk, and Kwok-wing Chau. "Flood prediction using machine learning models: Literature review." Water 10.11 (2018): 1536.

[18] Young, Chih-Chieh, Wen-Cheng Liu, and Chung-En Chung. "Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge." Neural Computing and Applications 26.7 (2015): 1631-1643.

[19] Nourani, Vahid, et al. "Applications of hybrid wavelet–artificial intelligence models in hydrology: a review." Journal of Hydrology 514 (2014): 358-377.