IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper | Vol.-7, Issue-4, April 2019 | E-ISSN: 2347-2693

Eyeopen for Predicting of Flood Flow by Statistical and Machine Learning Model

D.Suganya[1*], V.Savithri[2]

[1] Dept. of Computer Science, Women's Christian College, University of Madras, Chennai, India
[2]Dept. of Science and Technology, Women's Christian College, University of Madras, Chennai, India

*Corresponding Author: sugandilli@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i4.564568 | Available online at: www.ijcseonline.org

Accepted: 07/Apr/2019, Published: 30/Apr/2019

***Abstract*—** Floods are rare and dangerous disaster in minimum duration, which have the most destructive impact within urban and rural areas. The research on the advancement of flood prediction models contributed to risk reduction, to prevent the loss of human life, and reduction the property damage in floods. This paper structure the machine learning models, two separate models based on including and excluding the river flow were developed for each variable to quantify the importance of the river flow on the accuracy of the flood forecasting models. In this paper, aims to discovering more accurate and efficient prediction models. In recent research, two main approaches are developed in hydrological forecasting. The first approach is based on mathematical modeling. It models the physical dynamics between the principal components of the hydrological system. The second approach is based on modeling the statistical relationship between the hydrologic input and output, without explicitly considering the relationships that exist among the involved physical processes. The water level flow is deducted in this research, river flow proves the most and least improvement on the efficiency of the models applied for flood forecasting. As a result, this paper describes the most promising prediction methods for both long-term and short-term floods. This paper can be used as a predicting the flood by choosing the proper Machine Learning (ML) algorithm such as Support Vector Machine(SVM) and Artificial Neural Network(ANN) algorithm for showing higher accuracy.

***Keywords*—**flood prediction, hydrological model, machine learning, flood prediction, artificial intelligence, time series prediction.

I. Introduction

Flood prediction models are importance for hazard assessment and extreme event management. The floods caused big damage on agricultural and urban areas, especially cultural plants, houses, road railways. The rapid and accurate maps of affected areas are crucial for effective of rescue team. The procedure assures fast and reliable mapping of flooded areas using machine learning techniques. There are several steps in the mapping procedure, the most important one being input data preparation for machine learning. The output of the processing is classification model for detection of flooded areas. The importance of advanced systems in short-term and long-term prediction for flood and other hydrological events is strongly emphasized to alleviate damage. Today's major flood prediction models are mainly data-specific and include various simplified assumptions.

The mathematical expressions of physical processes and basin behaviour, such models benefit from specific techniques e.g., event-driven, deterministic, continuous, and hybrids. Physically based models were long used to predict hydrological events, such as storm rainfall/runoff shallow water condition hydraulic models of flow. The statistical models of autoregressive moving average (ARMA), multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) are the most common flood frequency analysis methods for modelling flood prediction. The drawbacks of the physically based and statistical models encourage the usage of advanced data-driven models, e.g., machine learning. Machine learning is a field of artificial intelligence (AI) used to induce regularities and patterns, providing easier implementation with low computation cost, as well as fast training, validation, testing, evaluation, with high performance compared to physical models, and less complexity. In performance of a number of physical and machine learning prediction models, showed a higher accuracy of machine learning models. In comparison to statistical models, machine learning models were used for prediction with greater accuracy. Many machine learning algorithms, e.g., artificial neural networks (ANNs), neuro-fuzzy, support vector machine (SVM), and support vector regression (SVR), were reported as effective for both short-term and long-term flood forecast. Machine learning algorithms have important characteristics that need to be carefully taken into consideration. Finally the effective algorithm to build the classification model and the most influential attributes for flood detection were defined.

II. LITERATURE REVIEW

In [1], Active Learning Method (ALM) is compared with Support Vector Machine (SVM) to predict long term simulation of daily stream flow in river. The daily discharge data were utilized for training and testing of the models. It result SVM is a well known method for run-off simulation and its capabilities have been demonstrated.

In [2], aimed to forecast the River Nile in Sudan using an Artificial Neural Network (ANN). This model validates the accuracy against the actual flow. This analysis indicates that ANN provides a reliable flood detecting in the River Nile.

In [3], authors explore of machine learning methods for flood forecasting in river. This analysis based on several machine learning algorithm and set of upstream and downstream flood was tested. It results Bayesian Linear model used for forecasting of extreme flood events.

In [4], authors present the knowledge of current and emerging trends in Artificial Neural Network (ANN) application research. This paper proposed feed forward and feedback propagation ANN model for research based on data analysis like accuracy, performance, volume, convergence and processing speed.

In [5], authors have proposed a two-level approach for clustering large data set for rainfall data prediction with Support Vector Machine (SVM). In this approach perform well direct clustering and to reduce the computation time.

In [6], authors have proposed Artificial Neural Network (ANN) is component to improve rainfall forecast. These techniques were used to provide more monthly rainfall forecast and it has an application where there is temperature and rainfall data.

In [7], authors propose a deep learning approach by integrating Stacked Auto Encoders (SAE) and Back Propagation Neural Network (BPNN) for the prediction of stream flow. This experiment research results in SAE-BP integrated algorithm perform much better.

In [8], Artificial Neural Network (ANN) model were applied in River to forecast the flood. The observations of data were river-stage and rainfall is validated since the river-stage observation has started. The input and output data of the neural network in river-stage predicted that it has very good accuracy.

In [9], authors describe Adaptive Neuro Fuzzy Inference System (ANFIS) and Gaussian Process Regression (GPR) for flood forecasting in River Cauvery in India. The technique were used for accuracy by the help of Mean Square Error (MSE) and the coefficient of correlation(R). It results in comparison of ANFIS and GPR to predict flood.

In [10], authors present an overview of flood forecasting in Internet of Things. This model is for Remote Monitoring and Controlling of Dams for observing and controlling water levels.

III. DATA SET

The flood prediction can be classified according to flood resource variables, i.e., water level, river flood, rainfall–discharge, river inflow, peak flow, river flow, rainfall–runoff, flash flood, rainfall, stream flow, seasonal stream flow, flood peak discharge, urban flood, plain flood, groundwater level, flood frequency analysis, extreme flow, typhoon rainfall, and daily flows. The first is that they are as good as their training, where the system learns the target task based on past data. If the data is scarce or does not cover varieties of the task, their learning falls short and they cannot perform well when they are put into work. The second aspect is the capability of each machine learning algorithm, which may vary across different types of tasks. This can be called a generalization problem, which indicates how well the trained system can predict cases it was not trained for, i.e., whether it can predict beyond the range of the training dataset. The dataset are traditionally rainfall and water level, measured either by ground rain gauges, or relatively new remote-sensing technologies such as satellites, multisensory systems, and radars. Each individual set of data undergo training, validation, verification, and testing. The principle behind the machine learning model workflow and the strategy for flood modelling improves better prediction.

IV. EXISTING SYSTEM MACHINE LEARNING METHODS IN FLOOD PREDICTION

The major machine learning algorithms applied to flood prediction include Artificial Neural Networks (ANNs) and Support Vector Machines (SVM). In the following, a brief description and background of these fundamental machine learning algorithms are presented.

A. Artificial Neural Networks (ANNs)

ANNs are efficient in mathematical modelling systems with efficient parallel processing, enabling them to mimic the biological neural network using inter-connected neuron units.

Feed forward neural networks (here called ANN) are a common type of artificial neural networks which applied in this study. An ordinary ANN model consists of input, hidden and output layers in which each layer has its nodes/neurons. Each layer is connected to the following layer via nodes. The nodes in the input layer which representing the input variables are transformed to hidden layer with weighted connections. The computations and processes are carried out in the hidden layer and then the nodes of the hidden layer are connecting to the output layer. Based on the relative importance of each input variable, appropriate weights between the connections of the nodes in the layer with those of the following layer are assigned.

ANNs are the most popular learning algorithms; it is efficient in modelling complex flood processes and accurate approximation. ANN as one of the most suitable modelling techniques which provide an acceptable generalization ability and speed compared to be the most conventional models. Despite the advantages of ANNs, there are a number of drawbacks associated with using ANNs in flood modelling, e.g., network architecture, data handling, and physical interpretation of the modelled system. A major drawback when using ANNs is the relatively low accuracy and the slow response in learning processes.

B. Support Vector Machine(SVM)

Support Vector Machines (SVMs) are a common type of data mining technique for classifying and regression purposes. They employ a hyper plane to separate data points. SVR is a type of support vector machine dealing with regression problems. SVM and SVR emerged as alternative machine learning methods to ANNs, with high popularity among hydrologists for flood prediction.SVM is greatly popular in flood modelling; it is a supervised learning which works based on the statistical learning theory and the structural risk minimization rule. The training algorithm of SVM builds models that assign new non-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 algorithm on training from past data. SVMs are both suitable for linear and nonlinear classification, it is efficient in mapping of inputs into feature spaces. The high computation cost of using support vector machine and their unrealistic outputs might be demanding, due to their heuristic and semi-black-box nature, the least-square support vector machine (LS-SVM) highly improved performance with acceptable computational efficiency. The alternative approach of the least-square support vector machine involves solving a set of linear tasks instead of complex quadratic problems. Nevertheless, there are still a number of drawbacks that exist, especially in the methods of application of seasonal flow prediction using LS-SVM.

V. PROPSED SYSTEM CLASSIFICATION OF MACHINE LEARNING (ML) METHODS

The most popular Machine learning model methods for flood prediction were identified in ANN and SVM. Considering the machine learning methods for application to floods, it is apparent that ANNs, SVMs are most popular. These machine learning methods can be categorized as single and hybrid model. The strategies involved developing hybrid machine learning models using soft computing techniques, statistical methods, and physical models rather than individual machine learning approaches. In hydrology, the definitions of short-term and long-term in learning technique the different phenomena vary. Short-term predictions for floods refer to hourly, daily, and weekly predictions, and they are used as warning systems. On the other hand, long-term predictions are mostly used for analysis purposes. Further, if the prediction leading time to flood is three days longer than the confluence time, the prediction is considered to be long-term. In the following, figure shows machine learning methods for flood prediction.

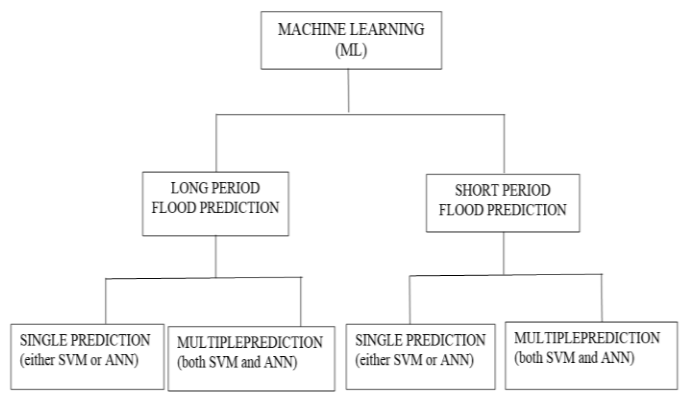

Table: Flood Prediction using Machine Learning Technique

Machine Learning Technique	Long Period Prediction		Short Period Prediction	
	SVM	ANN	SVM	ANN
Single Prediction	YES	NO	NO	YES
Multiple Prediction	YES	YES	YES	YES

A. Flood prediction in short period

Short-period lead-time flood predictions are considered to be the important research challenges, particularly in highly urban areas, for timely warnings to residences to reduce damage. This section is divided into two subsections—single and multiple methods of machine learning to individually investigate each group of methods.

1) Flood Prediction in short period using Single Method

ANN model to improve flood forecasting short-period flood prediction time through consideration of atmospheric conditions. They used satellite data from the dataset. This dataset includes hourly rainfall from rain gauges within the region. The ANN was reported to be considering a more accurate prediction than the statistical models. An analysis of single ML models for the prediction of short-period floods, considering the complexity of the algorithm, ease of use, running speed, accuracy, and input dataset. The quality of ML model prediction, continuously improved through using ensembles of ML methods, hybridization of ML methods, optimization algorithms, and/or soft computing techniques.

2) Flood prediction in short period using Multiple Methods

The Adaptive Neuro Fuzzy Inference System (ANFIS) hybrid model tuned by SVR provided superior prediction accuracy and good cost-effective computation for nonlinear and real-time flood prediction. The model with human interaction could provide better performance to predict the flood. An ANFIS model based on a dataset, which provided reliable hourly predictions. To improve the quality of prediction, in terms of accuracy, generalization, uncertainty, longer lead time, speed, and computation costs, there is an increase in building hybrid ML methods. This model had a fast development time, which also provided probabilistic forecasts to deal with uncertainties in prediction. This prediction system was reported as most useful and robust.

B. Flood prediction in long period

Long-period flood prediction is an importance for increasing knowledge and water resource management over longer periods of time, from weekly to monthly and annual predictions. This section is divided into two subsections—single and multiple methods of machine learning.

1) Flood prediction in long period using Single Method

SVM model for estimating stream flow and reservoir inflow for a long lead time for prediction. The prediction models were built using monthly river flow discharges for training, and for testing. AN analysis of single ML models for the prediction of long-period floods considering the algorithm, speed, accuracy, and input dataset. Thus builds a single ML methods long term prediction with high dataset in one method.

TABLE-1 Correlations Long Term

		rainfall	water flow	waterlevel
rainfall	Pearson Correlation	1	.987	1.000**
	Sig. (1-tailed)		.052	.010
	N	3	3	3
waterflow	Pearson Correlation	.987	1	.991*
	Sig. (1-tailed)	.052		.042
	N	3	3	3
	Pearson Correlation	1.000**	.991*	1
	Sig. (1-tailed)	.010	.042	
	N	3	3	3

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

2) Flood prediction in long period using Multiple Method

Data decomposition methods contributed highly to developing multiple methods for longer prediction lead time, good stability, great representativeness, and higher accuracy. These data decomposition methods were integrated with ANNs and SVM and they are expected to gain more popularity. This improves in prediction of accuracy and generalization capability. In fact, recent ensemble methods contributed to good improvements in speed, accurate prediction, and generalization. The most significant hybrid models were ensemble prediction models suitable for monthly prediction.

TABLE 2 Correlations Short Term

		rainfall	waterflow	waterlevel
rainfall	Pearson Correlation	1	.987	1.000**
	Sig. (1-tailed)		.051	.006
	N	3	3	3
waterflow	Pearson Correlation	.987	1	.990*
	Sig. (1-tailed)	.051		.045
	N	3	3	3
waterlevel	Pearson Correlation	1.000**	.990*	1
	Sig. (1-tailed)	.006	.045	
	N	3	3	3

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

VI. CONCLUSION

The state of ML modelling for flood prediction is advancement. This paper presents an overview of machine learning models used in flood prediction, and develops methods. The performance and accuracy of at least two machine learning models were compared ANN with SVM. SVM predict much better and efficient for flood flow prediction. The prediction models were classified into two categories according to short and long period prediction lead time, and further divided into categories of hybrid and single methods. The statistical measurement were used to analyse the prediction by using the correlation method, it predict in long period and short period model where SVM is efficiency rather than ANN algorithm. This analysis results in predicting waterfall and water level for the rain fall. This study helps to predict the extent of water flow range and the level of water flow where it cause flood in living area. The use of algorithm to improve the quality of machine learning model.

REFERENCES

[1] H. Shahraiyni, M. Ghafouri, S. Shouraki, B. Saghafian, and M. Nasseri, "Comparison between active learning method and support vector machine for runoff modeling," *J. Hydrol. Hydromechanics*, vol. 60, no. 1, pp. 16–32, 2012.

[2] S. H. Elsafi, "Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan," *Alexandria Eng. J.*, vol. 53, no. 3, pp. 655–662, 2014.

[3] J. Noymanee, N. O. Nikitin, and A. V. Kalyuzhnaya, "Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin," *Procedia Comput. Sci.*, vol. 119, no. 2017, pp. 288–297, 2017.

[4] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, 2018.

[5] R. UshaRani, T. K. R. Krishna Rao, and R. Kiran Kumar Reddy, "An Efficient Machine Learning Regression Model for Rainfall Prediction," *Int. J. Comput. Appl.*, vol. 115, no. 23, pp. 24–30, 2015.

[6] J. Abbot and J. Marohasy, "Forecasting of Medium-term Rainfall Using Artificial Neural Networks: Case Studies from Eastern Australia," *Eng. Math. Top. Rainfall*, 2018.

[7] F. Liu, F. Xu, and S. Yang, "A Flood Forecasting Model Based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network," *Proc. - 2017 IEEE 3rd Int. Conf. Multimed. Big Data, BigMM 2017*, pp. 58–61, 2017.

[8] M. Hitokoto and M. Sakuraba, "Applicability of the Deep Learning Flood Forecast Model Against the Inexperienced Magnitude of Flood," vol. 3, pp. 901–893, 2018.

[9] P. Misra and S. Shukla, "Implementation of Neuro-Fuzzy and Statistical Technique for Flood Forecasting in Cauvery Basin , India," no. 6, 2018.

[10] S. Vandana, D. D. Raj, B. Rushika, and V. M. Venkatesh, "Open Access," no. 11, pp. 444–447, 2018.

Authors Profile

Ms.D.Suganya pursed Bachelor of Computer Application from University of Madras, Chennai in 2014 and Masterof Science from University of Madras in year 2017. She is currently pursuing M.Phil(Computer Science)in Department of Computer Sciences, Women's Christian College, University of Madras, Chennai. Her main research work focuses on Machine Learning, Artificial Intelligence.

Mrs.V.Savithri is currently working as Assistant Professor in Department of Computer Science and technology, Women's Christian College, Chennai. She has completed Ph.d in Computer Science and her research focuses on Artificial Neural Network, Machine Learning and Artificial Intelligence. She published Research paper and she has attended many conferences.

