

# Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review

Nur-Adib Maspo<sup>\*1</sup>, Aizul Nahar bin Harun<sup>2</sup>, Masafumi Goto<sup>1</sup>, Faizah Cheros<sup>1</sup>, Nuzul Azam Haron<sup>3</sup>, and Mohd Nasrun Mohd Nawi<sup>4</sup>

<sup>1</sup> Department of Environmental Engineering and Green Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia. 54100 Kuala Lumpur, Malaysia

<sup>2</sup> Department of Management of Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia. 54100 Kuala Lumpur, Malaysia

<sup>3</sup> Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia

<sup>4</sup> School of Technology Management and Logistic, Universiti Utara Malaysia

\*Corresponding author e-mail: nuradibmaspo@gmail.com

**Abstract.** Flood disaster is a major disaster that frequently happens globally, it brings serious impacts to lives, property, infrastructure and environment. To stop flooding seems to be difficult but to prevent from serious damages that caused by flood is possible. Thus, implementing flood prediction could help in flood preparation and possibly to reduce the impact of flooding. This study aims to evaluate the existing machine learning (ML) approaches for flood prediction as well as evaluate parameters used for predicting flood, the evaluation is based on the review of previous research articles. In order to achieve the aim, this study is in two-fold; the first part is to identify flood prediction approaches specifically using ML methods and the second part is to identify flood prediction parameters that have been used as input parameters for flood prediction model. The main contribution of this paper is to determine the most recent ML techniques in flood prediction and identify the notable parameters used as model input so that researchers and/or flood managers can refer to the prediction results as the guideline in considering ML method for early flood prediction.

## 1. Introduction

Flood is the most devastating disaster that brings great damages such as loss of lives, and destruction of infrastructures, properties and environments, which consequently result in economic losses. These losses can be prevented or reduced by implementing non-structural measurement such as flood prediction so that the next coming flood can be forecast and flood information is available ahead of the event. By having the predicted flood event, governmental authority and people at the potential flooding area will be able to prepare for the action to be taken and having proper decision. In this case, lives, properties, infrastructures and environments can be protected and the effects on economic damage can be mitigated.

Research in the field of flood prediction has been conducted for a few decades to predict flood events and it continues to remain the challenging topic to date. In general, two types of flood prediction approaches are applied in flood prediction scenarios, one is physical principle-based models [1] that



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

include models based on the principles of physical processes such as rainfall-runoff model[2, 3], hydrodynamic model [4], soil and water assessment tools[5]. Although previous studies have proved that physical principle-based models has great ability to predict flood in different scenarios but the model is complex and taking long time in computation. Additionally, it also requires huge amount of input parameters that describe the physical characteristics, but the required data is not always available and not easy to collect. Furthermore, applying physical model requires comprehensive understanding and skill competence in hydrology as well as the ability to compute model with complexities. With these limitations of physical model, a data-driven model is an alternative model that has been used in flood prediction; among the typical models, autoregressive moving average (ARMA)[6], multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA)[7] are considered as traditional statistic models that are often used for flood frequency analysis(FFA) technique applied in forecasting flood. In comparison of physical principle-based models and statistical approaches, it is found that statistical approaches are more efficient in terms of computational cost and generalization, furthermore, more components are required to process physically based model. However, traditional statistical methods are reported to be less accurate in predicting flood and it is not suitable when applied for short-term flood prediction[8].

Advance data driven model such as machine learning (ML) is an attractive approach to overcome the limitation of physical principle-based models model and traditional statistical model. ML is emerging as a popular topic in hydrological problems specifically in time series flood forecasting. It is mainly to identify the relationship or pattern between the input parameters and the output. On top of that, ML models formulate the nonlinearity of flood into mathematical expression based on historical data without the need of fundamental knowledge in physical process. Another reason that makes ML models become popular is less computational cost; the process of ML is easy to implement and develop such as in model training, testing and evaluation, and with somewhat less complexity. Mosavi, Rabczuk and Varkonyi-Koczy [9] state that ML method is suitable for applying in flood prediction and its performance outperforms the conventional approaches and has been proved for greater accuracy.

Previous studies have applied diverse methods and techniques either within ML's boundary or integrated with other approaches in order to optimize the prediction accuracy. Accuracy measurement is used to evaluate model performance that researchers always evaluate through mean error (ME), mean squared error (MSE), root mean square error (RMSE), mean percentage error (MPE), mean absolute percentage error (MAPE) and squared correlation factor,  $R^2$ , which is also known as the correlation coefficient (CC). The higher the result accuracy is, the better the flood prediction model will be.

A few researches have been done on the comparison of flood prediction approaches but none has investigated and evaluated flood prediction approaches based on ML, and highlighted on the key parameters. Hence, this study intends to investigate and evaluate ML techniques and algorithms used in predicting flood as well as investigate the input parameters used for flood modelling. The outcome of this study will contribute to researcher or hydrologist in selecting the suitable ML techniques for implementing flood prediction and analysis task. Additionally, this study is done through systematic literature review that is further discussed in the next section.

## 2. Related work

There are a few researches that were similarly conducted reviewing the methods and techniques used in flood prediction, and analysis such as by Devia, Ganasri and Dwarakish [1], who had reviewed hydrological model and discussed on the three types of flood prediction models. They are physical principle-based model, conceptual model and data driven model. Advantages and drawbacks of each model were discussed. However, this paper focused on physical principle-based models that the authors stressed on different types and applications of physically based model such as SWAT, SHE/MIKE SHE model, but data driven models are not discussed in depth. Teng, Jakeman, Vaze, Croke, Dutta and Kim

[4] have reviewed and compared different types of hydrological model and hydrodynamic model, and further categorized them into 1D model, 2D model and 3D model. Based on the revision of research articles in recent years, there is only one paper that specifically reviewing data driven model based on ML technique [8], the study was done through comparative study of the most recent ML techniques for flood prediction model but no discussion on the important of input parameters.

### **3. Methodology**

This study is conducted through systematic literature review, online database such as Scopus was the main database for article extraction. The period was scoped for the most recent 5 years from year 2014 to 2018. The terms used in searching for related articles are “flood prediction” and “flood forecasting”. As a result, the total of 84 articles were extracted, and 29 out of 84 were selected as the relevance articles.

This article is structured into five parts; Part 1 presents introduction and background of this study. Part 2 presents related studies. Part 3 presents the popular ML techniques applying in flood prediction. The review result and discussion were explained in part 4 and the conclusion of this article is stated in the last section.

### **4. Machine learning technique and algorithms apply in flood prediction scenarios.**

Machine learning (ML) is defined by Samuel [10] as “Field of study that gives computers the ability to learn without being explicitly programmed”, which basically means that in ML study we can assign some tasks to the computer and let it learn based on the task given, and we do not need to obviously code the program. Another well elaborated by Mitchell [11] “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E”. to relate this quote in applying ML to flood prediction, we can say that to predict the future flood events (Task T), we can process through machine learning algorithms with available historical data (Experience E) and when it is effectively “learned”, then the model will predict better for future flood events with better performance measurement (P).

#### *4.1. Types of machine learning*

Different sources describe different types of learning that categorize under ML method. There are 3 types of learning that commonly applied in machine learning. They are supervised learning, unsupervised learning and reinforcement learning[12-14]

##### *4.1.1 Supervised Learning*

Supervised learning is the type of ML that needed to train with labelled data, this leaning step requires a pair of input data and preferred output. Basically, we teach the computer how to do the procedure, then let it use its new knowledge to do the next step. This is also called learning from exemplars. Some techniques that are categorized under supervised learning support vector machine (SVM), Naïve Bayes, deep learning etc. The recent applications that are applied in supervised learning are spam filter for receiving email, Cancer detection that is applied in healthcare and detecting fraud activities of credit card usage.

##### *4.1.2 Unsupervised Learning*

Unsupervised learning is another type of ML that does not need to train with labelled data and the learning process is only provides inputs without desired output. The algorithms will keep trying to explore the input data and find the pattern that similar and group as the same category. In another word, we allow the computer to learn how to carry out the task, and use this to determine structure and patterns in data. Some of unsupervised techniques include neural network, clustering, and K-nearest neighbour (KNN), etc. Examples of recent application that is by unsupervised learning is customer segmentation into groups with the same preferences, detecting unusual access to a website.

#### 4.1.3. Reinforcement Learning

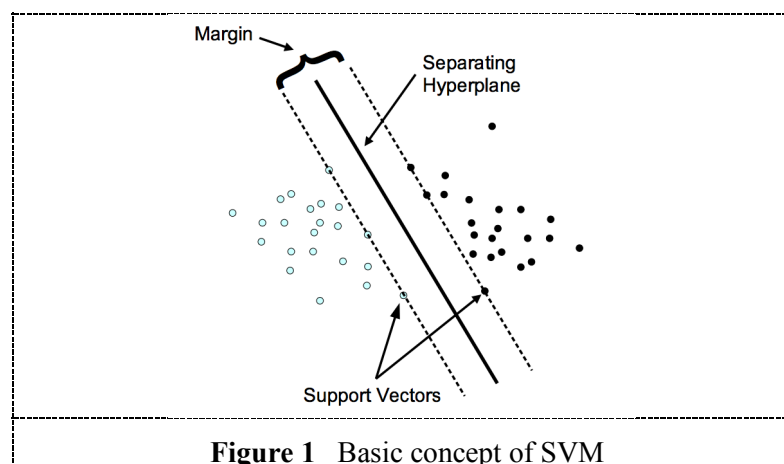
This type of ML is classified between supervised and unsupervised learning. This ML works as the algorithm gets informed when the output displays incorrectly, whereas the learning algorithm will not get informed about the way to correct it. To get the right output, the learning algorithm needs to investigate and attempt different possibilities.

In order to predict the flood, selecting and applying algorithms is required in the process of ML. The following section will further explain different types of ML algorithms for the application of flood prediction.

#### 4.2. Type of machine learning algorithms

##### 4.2.1. Support Vector Machine (SVM)

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. Figure 1 shows the basic concept of SVM. SVM is one of ML algorithms that is considered popular in modelling flood, it is also known as robust and efficient algorithm in predicting flood [15, 16]. Recent researches have applied SVM in modelling flood [17-19]. Li, Ma, Jin and Zhu [20] implemented a novel flood forecasting model based on SVM and boosting learning algorithms, and found that the overall accuracy is 0.983 which is close to 1, hence the prediction precision is high. On top of that this study applied boosting algorithm in order to increase accuracy.



**Figure 1** Basic concept of SVM

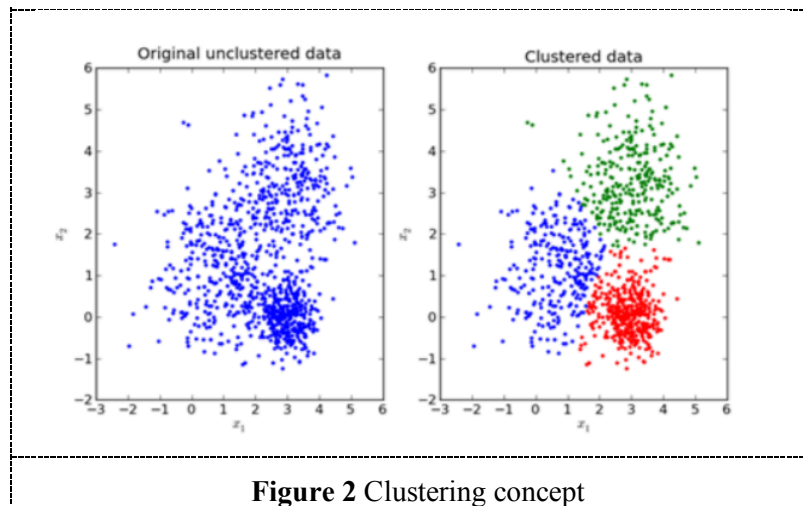
##### 4.2.2. Bayesian algorithm

Bayesian algorithm is generally used for solving classification problem [21]. The principal design of Naïve Bayes depends on the conditional probability. Naive Bayes is a probabilistic ML algorithm that can be expanded broadly in classification problem i.e. classifying spam in email, documents classification, sensation estimation etc. In flood prediction scenarios Jangyodsuk, Seo, Elmasri and Gao [22] used Bayesian algorithm with an optimization function to maximize mutual information. Noymanee, Nikitin and Kalyuzhnaya [23] proposed Bayesian linear model for predicting flood in Pattani province, Thailand. This algorithm was applied for rebuild the historical rivers floods and predicting the possible flood.

##### 4.2.3. Clustering

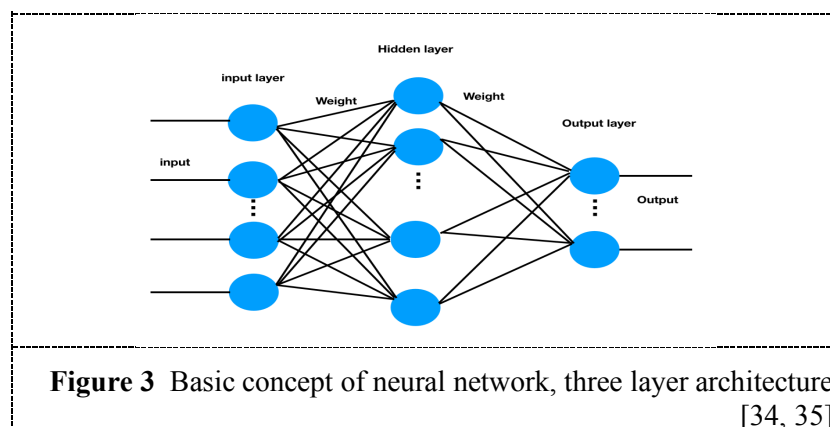
Clustering known as grouping is classified under unsupervised learning, when the algorithm begins, it automatically forms grouping. The class that hold the same characteristics is categorized in the same

group. Dissimilarity will be grouped as different clusters. K-mean is an algorithm under clustering technique that useful with analysis of large dataset. Many researches applied clustering in variety of hydrological conditions like modelling of rainfall-runoff [24], flood estimation for ungauged catchment [25], flood risk assessment [26] etc. Sood, Sandhu, Singla and Chang [27] used K-mean clustering algorithm to estimate the recent flood situation and evaluate flood inundation area. This study integrated ML with Holt-Winter's forecasting method for flood prediction by utilizing meteorological data as input dataset. The basic concept of clustering is shown in Figure2.



#### 4.2.4. Artificial Neural Network (ANN)

Artificial neural network is defined as the connection of input signal and output signal utilizing a model that imitates network of a biological brain which responds to any incentives from sensory inputs. ANN applies a network of artificial neurons or nodes to produce the output e.g. input layer connected to the hidden layer with the assigned weighting of each connection, the output that propagated from the hidden with the assign weighting will produce the output as shows in Figure 3, the basic concept of neural network. In flood forecasting, ANN has attracted the attention of researchers enduringly [28] as it has the great competence in nonlinear modelling and complex framework without clear physical justification. Hydrologist examines ANN in different flood scenarios such as flood inundation forecasting model [29], rainfall-runoff analysis[30], stream flow forecasting [31]. Recent researches have introduced hybridization approach within ML model or other models like physical based model in order to maximize accuracy rate[32, 33]. The components of NN such as Deep learning, BPNN and FFNN are further explain below.



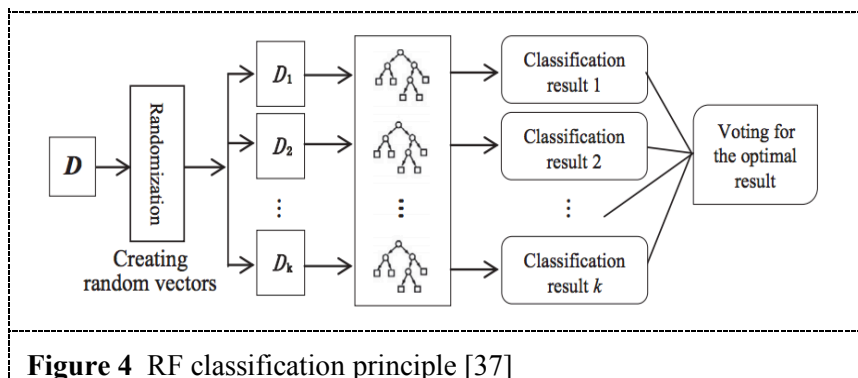
- Deep Learning is the powerful technique of learning artificial neural network, it uses multiple layers to increasingly extract higher level features from raw input. Deep learning is the best solution to solve problem in image recognition, speech recognition, and natural language processing. It is also best solution in time series data analysis that can be applied in flood prediction problem.
- Back Propagation Neural Network (BPNN) is the algorithm that processed within artificial neural network. When feed forward neural network is processed, if error found back propagation algorithm will take part and adjust the weighting then process again till the result is satisfied.
- Feed Forward Neural Network is the algorithm used in neural network where the information direction is fed forward i.e. from input layer to hidden later then output later, however, weightings are not adjusted.

#### 4.2.5. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an algorithm used for dimension reduction of dataset when required, whereas the variability of the dataset has to be maintained as much as possible[36]. This technique is used in data exploration and preparation in ML method.

#### 4.2.6. Random Forest

Random Forest (RF) is a learning algorithm that works as a large collection of uncorrelated decision tree then a lot of decision trees are created and use them to make a classification based on the most voted class for result optimization. The following Figure 4 shows random forest classification principle. In hydrological, RF has been used to develop flood hazard risk model [37]. Garcia, Retamar and Javier [38] used RF algorithm for modelling flood forecasting and deliver flood advisory to the users beforehand.



**Figure 4** RF classification principle [37]

## 5. Result and discussion

ML method is popular in the field of hydrology. Researchers conducted research by applying different types of ML techniques and algorithms for flood prediction in order to maximize prediction accuracy and model optimization. According to the survey as summarized in Table 1, shows the summary of ML techniques, strategies and input parameters used in different flood prediction scenarios. Out of 29 papers, there are 4 papers that used single ML algorithm i.e. ANN[39, 40], Bayesian algorithm[22, 23] WEKA data mining technique[41]. 10 papers out of 29 applied integrated approach with other models, the integration includes within the boundary of ML method or integrated ML with physically based model i.e. HEC-HMS. ANN is considered as the most widely used machine learning algorithm in flood prediction, with the add-on algorithm such as BPN, FFNN and MLP will contribute to model optimization, in this survey there are 11 papers of the total 29 paper categorized as algorithm optimization study. 5 papers out of 29 papers used the collection of related algorithms to conduct their study. SVM integrated with other algorithms such as GRA, boosting and PCA[20, 42]. Deep learning is another method that recently being use in flood prediction scenarios and it obtains high accuracy rate.

Data mining tool such as WEKA and other physically based model such as TOPMODEL, Xinanjiang model, MIKEFLOOD are integrating with ML model for the improvement of model accuracy.

Mosavi, Ozturk and Chau [8] suggested four strategies for improving flood prediction accuracy, first strategy is hybridization that is to integrate two or more within ML methods or to integrate ML methods with physically based model. Second strategy is algorithms ensemble which the model collected and run multiple related algorithms and choose the best algorithm that obtain the most accurate result. Using ensemble technique will decrease the uncertainty of prediction. Third is algorithm optimization is another strategy to enhance the quality of ML algorithms, e.g. to improve ANN, applying BPN and FFMLP for model optimization. Lastly, data decomposition is another way to improve prediction accuracy. When the quality of dataset is improved, the prediction accuracy will parallelly improved. Hence, this study found that model optimization and hybridization within ML approach have gained much popularity as they often perform better than individual models [43]. On the other hand, hybridization of ML approach with physically based model remain less popular as compare to hybridization within ML approach, ensemble and model optimization. The reason could be the cost of computational and model complexity may contribute to less popularity.

On top of that, input parameters are important elements in hydrological studies. Chen and Han [44] classified hydrological data in three dimensions, one is natural dimension that included measurements of precipitation, stream flow, soil moisture, ground water, temperature and humidity, etc. The methods of collecting data can be from in-situ observation or point to radar application, remote sensing, satellites, and drone. Second is social dimension which refers to the reaction of human society towards of water environment. The data includes Twitter data that is now popular and able to capture, analyse and create meaningful information in water environment applications. Another dimension is business dimension that covered data of water supply, waste water collection and treatment, etc. However, this dimension is not directly related to flood management, but is it more into water management.

In the 29 reviewed articles, natural data source is the only dimension being used, none has considered social and business dimensions. Water level and rainfall remain the dominant input parameters that required in all flood modelling due to it is significantly contribute in the field of flood scenarios. Additionally, input parameters have to be precisely identified in flood modelling based on the selected conditions in order to determine the most accurate output. When prediction output generates unacceptable number of errors, the input parameters has to reset and re-process the model then obtain acceptable result. According to Devia, Ganasri and Dwarakish [1], the best model is the one which gives result closed to reality with the use of minimum parameters and less model complexity.

**Table 1.** shows the summary of ML techniques, strategies and input parameters used in different flood prediction scenarios

<b>ML techniques</b>	<b>Strategy for model improvement</b>	<b>Input parameters</b>	<b>References</b>
ANN + HS + DE	Algorithms ensemble	gauge, area, velocity, discharge, rainfall, average temperature, average wind speed and pressure	[45]
ANN + Gradient Descent + Levenberg Marquardt	Algorithms optimization	water level, humidity, pressure, rainfall,	[46]
Multi-layer Perceptron	Single algorithm	water level, rainfall	[39]

ANN + BPN + FFNN + LM	Algorithm optimization	water level	[47]
Clustering + Holt-Winter's	Hybridization	metrological data	[27]
ANN + MLP + LM + Fire fly algorithm	Hybridization	stream flow	[46, 48]
Deep learning + stacked autoencoders (SAE) + BPNN	Hybridization	rainfall	[49]
ANN + Gradient Descent(GD) + Levenberg Marquardt(LM) + Bayesian Regularization (BR)	Algorithms ensemble	temperature, humidity, pressure, wind, and level. rainfall, sea level, and water level.	[50]
Apriori algorithm + Data mining tools (WEKA)	Hybridization	flood area, water level, flood status	[41]
ANN	Single algorithm	water lever, rainfall	[40]
Neural network regression + Bayesian linear regression + Boosted decision tree regression + Decision Forest Regression + linear regression model	Algorithms ensemble	water level	[23]
NWP + BPN	Hybridization	global parameter for NWP model in PNG and JPEG format and local parameter for temperature, humidity, wind speed.	[32]
Datamining technique (WEKA)	Single algorithm	water level	[51]
ANN + Chaos theory	Algorithm optimization	water level and rainfall during flood	[52]
Bayesian based learning algorithm	Single algorithm	water level, rainfall and other hydrological data	[22]
ANN, feed forward Multilayer perceptron (FFMLP)	Algorithm optimization	water level during flood	[53]
ANN+ PCA + Fruit fry algorism	Algorithm optimization	water flow	[33]
Random forest algorithm	Algorithms ensemble	water level and rainfall	[38]



Integrated hydrological model and a symbolic regression method	LISFLOOD regression	Hybridization	DEM, water level, rainfall	[54]
SVM algorithm + principal analysis (KPCA)	+ boosting + kernel component	Algorithm optimization	stream low data	[20]
multilayer perceptron (MLP) + search (CS)	the Cuckoo	Algorithm optimization	water level data	[55]
ANN + BPN+ RFN		Algorithm optimization	Min, max and avg temperature, min, max and avg humidity, avg wind speed, avg sea level pressure and avg rainfall.	[28]
HEC-HMS + ANFIS	GANN +	Hybridization	rainfall and runoff discharge	[56]
ANN + Marquardt(LM) algorithm	+Levenberg-	Algorithm optimization	water level	[57]
SVM + GRA		Hybridization	Permeability, perforation, oil saturation, supply oil radius, perforation density, control area, control reserve, reservoir thickness, degree of reservoir drilling, water cut, hole radius, flowing bottom, hole pressure	[42]
Clustering + KNN		Algorithm optimization	rainfall	[58]
OLAP-based multidimensional cube+ MIKE FLOOD		Hybridization	rainfall and water level	[59]
ANN + Fuzzy Logic+ BPN		Algorithm optimization	river discharge data	[60]
Hybrid model of the Xinanjiang model and TOPMODEL		Hybridization	rainfall runoff depth	[61]

## 6. Conclusion

This study has compared and evaluated ML methods applied in flood prediction scenarios and parameters in the past 5 years. The findings show that ANN is the most popular ML method in predicting flood. However, most researchers applied model optimization strategy in order to improve the model accuracy by adding BPN and FFMLP into ANN model. Hybridization within ML method is another

technique that many researchers have conducted in their studies, however, hybridization of ML with physical principle-based model is less popular as compare to hybridization within ML technique. The less popularity of hybridization between physical principle-based model and ML could be due to the complexity in processing physical principle-based model. Model ensemble is another technique that used to stabilize the prediction accuracy but only few studies adopted ensemble technique. Hence, more opportunities are open for researchers to explore this technique in the area of flood prediction. It is suggested that the future research is needed to find out the reason of less popularity in model hybridization of ML integrated with physical principle-based model and include the study of data decomposition. Additionally, with regards to input parameters, this study only found the natural dimension that commonly used as input parameters for all the reviewed papers, it is recommended that social dimension should be included as data source to capture the most recent humans' thought toward water related condition as it may significantly contribute to improve the prediction accuracy. Furthermore, ML application in the field of environment and water management should be extended to various scenarios and considering multi sources data such as image, audio, video and text in processing and analysing data in order to obtain the useful knowledge for preparing, managing and preventing damages.

### Nomenclatures

ANN	Artificial neural network
MLR	Multiple linear regression
ML	Machine learning
ARIMA	Auto regressive integrated moving average
ARMA	Auto regressive moving average
SVM	Support vector machines
FFNN	Feed-forward neural network
MLP	Multilayer perceptron
BPNN	Backpropagation neural network
FFA	Flood frequency analysis
SWAT	Soil water analysis tool
KNN	K-nearest neighbor
NN	Neural network
PCA	Principle component analysis
RF	Random forest
HS	Harmony search
DE	Differential evaluation
LM	Levenberg– Marquardt
NWP	Numerical weather prediction
CS	Cuckoo search
HEC-HMS	Hydrological engineering center - hydrological modeling system
ANFIS	Adaptive neuro- fuzzy inference system
OLAP	Online analytical processing
GANN	Genetic algorithm neural network

### References

- [1] Devia G K, Ganasri B and Dwarakish G 2015 A review on hydrological models *Aquatic Procedia* **4** 1001-7
- [2] Piman T and Babel M 2012 Prediction of rainfall-runoff in an ungauged basin: case study in the mountainous region of northern Thailand *Journal of Hydrologic Engineering* **18** 285-96

- [3] Fernández-Pato J, Caviedes-Voullième D and García-Navarro P 2016 Rainfall/runoff simulation with 2D full shallow water equations: Sensitivity analysis and calibration of infiltration parameters *Journal of hydrology* **536** 496-513
- [4] Teng J, Jakeman A J, Vaze J, Croke B F, Dutta D and Kim S 2017 Flood inundation modelling: A review of methods, recent advances and uncertainty analysis *Environmental Modelling & Software* **90** 201-16
- [5] Narsimlu B, Gosain A K, Chahar B R, Singh S K and Srivastava P K 2015 SWAT model calibration and uncertainty analysis for streamflow prediction in the Kunwari River Basin, India, using sequential uncertainty fitting *Environmental Processes* **2** 79-95
- [6] Eni D and Adeyeye F J 2015 Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria *Journal of Geoscience and Environment Protection* **3** 91
- [7] Ab Razak N H, Aris A Z, Ramli M F, Looi L J and Juahir H 2018 Temporal flood incidence forecasting for Segamat River (Malaysia) using autoregressive integrated moving average modelling *Journal of Flood Risk Management* **11** S794-S804
- [8] Mosavi A, Ozturk P and Chau K-w 2018 Flood prediction using machine learning models: Literature review *Water* **10** 1536
- [9] Mosavi A, Rabczuk T and Varkonyi-Koczy A R 2017 Reviewing the novel machine learning tools for materials design. In: *International Conference on Global Research and Education*: Springer) pp 50-8
- [10] Samuel A L 1988 *Computer Games I*: Springer) pp 366-400
- [11] Mitchell T M 1999 Machine learning and data mining *Communications of the ACM* **42**
- [12] Qiu J, Wu Q, Ding G, Xu Y and Feng S 2016 A survey of machine learning for big data processing *EURASIP Journal on Advances in Signal Processing* **2016** 67
- [13] Alpaydin E 2009 *Introduction to machine learning*: MIT press)
- [14] Shalev-Shwartz S and Ben-David S 2014 *Understanding machine learning: From theory to algorithms*: Cambridge university press)
- [15] Dibike Y B, Velickov S, Solomatine D and Abbott M B 2001 Model induction with support vector machines: introduction and applications *Journal of Computing in Civil Engineering* **15** 208-16
- [16] Jajarmizadeh M, Lafdani E K, Harun S and Ahmadi A 2015 Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran *KSCE Journal of Civil Engineering* **19** 345-57
- [17] Tehrany M S, Pradhan B, Mansor S and Ahmad N 2015 Flood susceptibility assessment using GIS-based support vector machine model with different kernel types *Catena* **125** 91-101
- [18] Tehrany M S, Pradhan B and Jebur M N 2015 Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method *Stochastic environmental research and risk assessment* **29** 1149-65
- [19] Zhu S, Zhou J, Ye L and Meng C 2016 Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River, China *Environmental Earth Sciences* **75** 531
- [20] Li S, Ma K, Jin Z and Zhu Y 2016 A new flood forecasting model based on SVM and boosting learning algorithms. In: *2016 IEEE Congress on Evolutionary Computation (CEC)*: IEEE) pp 1343-8
- [21] Jiang L, Li C, Wang S and Zhang L 2016 Deep feature weighting for naive Bayes and its application to text classification *Engineering Applications of Artificial Intelligence* **52** 26-39
- [22] Jangyodsuk P, Seo D-J, Elmasri R and Gao J 2015 Flood prediction and mining influential spatial features on future flood with causal discovery. In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*: IEEE) pp 1462-9
- [23] Noymanee J, Nikitin N O and Kalyuzhnaya A V 2017 Urban Pluvial Flood Forecasting using Open Data with Machine Learning Techniques in Pattani Basin *Procedia computer science* **119** 288-97

- [24] Zahmatkesh Z, Karamouz M and Nazif S 2015 Uncertainty based modeling of rainfall-runoff: combined differential evolution adaptive metropolis (DREAM) and K-means clustering *Advances in water resources* **83** 405-20
- [25] Latt Z Z, Wittenberg H and Urban B 2015 Clustering hydrological homogeneous regions and neural network based index flood estimation for ungauged catchments: an example of the Chindwin River in Myanmar *Water resources management* **29** 913-28
- [26] Xu H, Ma C, Lian J, Xu K and Chaima E 2018 Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China *Journal of hydrology* **563** 975-86
- [27] Sood S K, Sandhu R, Singla K and Chang V 2018 IoT, big data and HPC based smart flood management framework *Sustainable Computing: Informatics and Systems* **20** 102-17
- [28] Chai S-S, Wong W K and Goh K L 2016 Backpropagation Vs. Radial Basis Function Neural Model: Rainfall Intensity Classification For Flood Prediction Using Meteorology Data *JCS* **12** 191-200
- [29] Chang L-C, Amin M, Yang S-N and Chang F-J 2018 Building ANN-based regional multi-step-ahead flood inundation forecast models *Water* **10** 1283
- [30] Chakravarti A, Joshi N and Panjiar H 2015 Rainfall Runoff Analysis Using the Artificial Neural Network *Indian Journal of Science and Technology* **8** 1-7
- [31] Kasiviswanathan K, He J, Sudheer K and Tay J-H 2016 Potential application of wavelet neural network ensemble to forecast streamflow for flood management *Journal of Hydrology* **536** 161-73
- [32] Šaur D 2017 Forecasting of convective precipitation through NWP models and algorithm of storms prediction. In: *Computer Science On-line Conference*: Springer) pp 125-36
- [33] El-Telbany M E 2016 Improving the Predictability of GRNN using fruit fly optimization and PCA: The Nile Flood Forecasting. In: *International Conference on Advanced Intelligent Systems and Informatics*: Springer) pp 310-9
- [34] Gurney K 2014 *An introduction to neural networks*: CRC press)
- [35] Tayfur G, Singh V, Moramarco T and Barbetta S 2018 Flood hydrograph prediction using machine learning methods *Water* **10** 968
- [36] Howley T, Madden M G, O'Connell M-L and Ryder A G 2005 The effect of principal component analysis on machine learning accuracy with high dimensional spectral data. In: *International Conference on Innovative Techniques and Applications of Artificial Intelligence*: Springer) pp 209-22
- [37] Wang Z, Lai C, Chen X, Yang B, Zhao S and Bai X 2015 Flood hazard risk assessment model based on random forest *Journal of Hydrology* **527** 1130-41
- [38] Garcia F C C, Retamar A E and Javier J C 2015 A real time urban flood monitoring system for metro Manila. In: *TENCON 2015-2015 IEEE Region 10 Conference*: IEEE) pp 1-5
- [39] Widiarsari I R and Nugroho L E 2017 Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. In: *2017 International Conference on Innovative and Creative Information Technology (ICITech)*: IEEE) pp 1-5
- [40] Bakar M A A, Aziz F A A, Hussein S F M, Abdullah S S and Ahmad F 2017 Flood Water Level Modeling and Prediction Using Radial Basis Function Neural Network: Case Study Kedah. In: *Asian Simulation Conference*: Springer) pp 225-34
- [41] Harun N A, Makhtar M, Aziz A A, Mohamad M and Zakaria Z A 2017 Implementation of Apriori Algorithm for a New Flood Area Prediction System *Advanced Science Letters* **23** 5419-22
- [42] Fan J and Li X 2015 Prediction of the productivity of steam flooding production wells using Gray Relation Analysis and Support Vector Machine *Journal of Computational Methods in Sciences and Engineering* **15** 499-506

- [43] Al-Jarrah O Y, Yoo P D, Muhaidat S, Karagiannidis G K and Taha K 2015 Efficient machine learning for big data: A review *Big Data Research* **2** 87-93
- [44] Chen Y and Han D 2016 Big data and hydroinformatics *Journal of Hydroinformatics* **18** 599-614
- [45] Panigrahi B K, Das S, Nath T K and Senapati M R 2018 An application of data mining techniques for flood forecasting: Application in rivers Daya and Bhargavi, India *Journal of The Institution of Engineers (India): Series B* **99** 331-42
- [46] Bande S and Shete V V 2017 Smart flood disaster prediction system using IoT & neural networks. In: *Smart Technologies For Smart Nation (SmartTechCon), 2017 International Conference On: IEEE* pp 189-94
- [47] Sahagun M A M, Cruz J C D and Garcia R G 2017 Wireless sensor nodes for flood forecasting using artificial neural network. In: *2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM): IEEE* pp 1-6
- [48] Khatibi R, Ghorbani M A and Pourhosseini F A 2017 Stream flow predictions using nature-inspired Firefly Algorithms and a Multiple Model strategy—Directions of innovation towards next generation practices *Advanced Engineering Informatics* **34** 80-9
- [49] Liu F, Xu F and Yang S 2017 A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network. In: *2017 IEEE third International conference on multimedia big data (BigMM): IEEE* pp 58-61
- [50] Keong K C, Mustafa M, Mohammad A J, Sulaiman M H and Abdullah N R H 2016 Artificial neural network flood prediction for sungai isap residence. In: *2016 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS): IEEE* pp 236-41
- [51] Harun N A, Makhtar M, Aziz A A, Zakaria Z A, Abdullah F S and Jusoh J A 2017 The application of apriori algorithm in predicting flood areas *International Journal on Advanced Science, Engineering and Information Technology* **7** 763-9
- [52] Furquim G, Pessin G, Faiçal B S, Mendiondo E M and Ueyama J 2016 Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory *Neural computing and applications* **27** 1129-41
- [53] Jaafar K, Ismail N, Tajjudin M, Adnan R and Rahiman M H F 2016 Hidden neuron variation in multi-layer perceptron for flood water level prediction at Kusial station. In: *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA): IEEE* pp 346-50
- [54] Merkurjeva G, Merkurjev Y, Sokolov B V, Potryasaev S, Zelentsov V A and Lektauers A 2015 Advanced river flood monitoring, modelling and forecasting *Journal of computational science* **10** 77-85
- [55] Phitakwinai S, Auephanwiriyaikul S and Theera-Umpon N 2016 Multilayer perceptron with Cuckoo search in water level prediction for flood forecasting. In: *2016 International Joint Conference on Neural Networks (IJCNN): IEEE* pp 519-24
- [56] Young C-C, Liu W-C and Chung C-E 2015 Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge *Neural Computing and Applications* **26** 1631-43
- [57] Wan D, Xiao Y, Zhang P and Leung H 2015 Hydrological big data prediction based on similarity search and improved BP neural network. In: *Big Data (BigData Congress), 2015 IEEE International Congress on: IEEE* pp 343-50
- [58] Di Y, Ding W, Mu Y, Small D L, Islam S and Chang N-B 2015 Developing machine learning tools for long-lead heavy precipitation prediction with multi-sensor data. In: *2015 IEEE 12th International Conference on Networking, Sensing and Control: IEEE* pp 63-8
- [59] Jo H-J, Lee Y-J, Seo J-H and Choi J-T 2015 Flood inundation prediction model on spatial characteristics with utilization of OLAP-based multidimensional cube information *International Journal of Software Engineering and Its Applications* **9** 1-10

- [60] Kulkarni N K and Shete V V 2014 Hybrid neuro-fuzzy approach for flood prediction and Dam gate control. In: *International Conference on Information Society (i-Society 2014)*: IEEE) pp 213-8
- [61] Liu S, Xu J, Zhao J, Xie X and Zhang W 2014 Efficiency enhancement of a process-based rainfall–runoff model using a new modified AdaBoost. RT technique *Applied Soft Computing* **23** 521-9

### **Acknowledgement**

Authors are highly thankful to Universiti Teknologi Malaysia for its continuous financial support.