Available online at http://ijdea.srbiau.ac.ir

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol. 9, No. 4, Year 2021 Article ID IJDEA-00422, Pages 43-90
Research Article

International Journal of Data Envelopment Analysis

Science and Research Branch (IAU)

# Application of machine learning models for flood risk assessment and producing map to identify flood prone areas: literature review

**P. Firoozishahmirzadi[*] [1], S. Rahimi [2], Z. Esmaeili Seraji[3]**

[1] Delft University of Technology, Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft, Netherlands,
[2] Department of Mathematics, University of Mazandaran, Iran.
[3] Department of healthcare, Islamic Azad University of Sari, Sari, Iran.

Received 3 May 2021, Accepted 8 October 2021

**Abstract**

Floods as the most destructive natural disaster are highly complex to model. The research on the advancement of flood risk assessment models contributed to risk reduction, policy suggestion, reduction of the property damage and minimization of the loss of human life. During the past two decades, machine learning methods contributed highly in the advancement of modeling systems, providing better performance and cost-effective solutions. Researchers through introducing novel ML methods and hybridizing of them aim at discovering more accurate and efficient models. The main contribution of this literature review is to demonstrate the state of ML models from two perspectives; 1-flood risk assessment, 2- producing flood reliable map to give insight into the most suitable models. In this literature is shown the important ML models that can have impressive effect on flood models are Super Vector Mane, Decision Tree, Logistic Regression and Random Forest respectively. Hybridization different kind of ML methods, data fusion that is a prevalent way to deal with imperfect raw data for capturing reliable, ensemble algorithm and model optimization are reported as the most effective strategies for the improvement of ML methods. Random Forest models do well with high dimensional data and their flexibility makes them suitable for solving more problems. ANN models are especially good at modeling multifarious nonlinear networks that are difficult to describe with functions directly.

**Keywords:** Flood Risk Assessment (FRA); hydrologic model; ensemble Machine Learning; Artificial Neural Networks (ANNs); Support Vector Machine (SVM); natural hazards & disasters; Decision Tree (DT); Artificial Intelligence (AI).

[*] Corresponding author: Email: Parisa.firoozi@gmail.com

## 1. Introduction

Floods are the most frequent type of natural disaster and occur when an overflow of water submerges land that is usually dry. Floods are often caused by heavy rainfall, rapid snowmelt or a storm surge from a tropical cyclone or tsunami in coastal areas. Floods can cause widespread devastation, resulting in loss of life and damages to personal property and critical public health infrastructure. Between1998-2020, floods affected more than 2 billion people worldwide. People who live in floodplains or non-resistant buildings, or lack warning systems and awareness of flooding hazard, are most vulnerable to floods. [1]. Also the global warming and climate changes determined a considerable increase in the frequency of floods and their related damages. Therefore, the high accuracy identification of flood susceptible areas plays a key role in flood warnings and risk reduction. Early notification of flood incident could benefit the authorities and public to devise both short and long terms preventive measures, to prepare evacuation and rescue mission, and to relieve the flood victims. Geographical locations of affected areas and respective severities, for instances, are among the key determinants in most flood administration. Thus far, an effective means of anticipating flood in advance remains lacking. Existing tools were typically based on manually input and prepared data. The processes were tedious and thus prohibitive for real-time and early forecasts. Furthermore, these tools did not fully exploit more comprehensive information available in current big data platforms. Mapping flood-prone areas is a key activity in flood disaster management. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, protection, and preparedness [1]. Flood prediction models are of significant importance for hazard assessment and extreme event management. Robust and accurate flood risk assessment contribute highly to water recourse management strategies, policy suggestions and analysis, and further evacuation modeling [2]. Thus, the importance of advanced systems for flood assessment, producing flood reliable map are strongly emphasized to alleviate damage. However, the prediction of flood lead time and occurrence location is fundamentally complex due to the dynamic nature of climate condition. Therefore, today's major flood prediction models are mainly data-specific and involve various simplified assumptions [3]. Thus, to mimic the complex mathematical expressions of physical processes and basin behavior, such models benefit from specific techniques e.g., event-driven, empirical black box, lumped and distributed, stochastic, deterministic, continuous, and hybrids [4]. Physically based models [5] were long used to predict hydrological events, such as storm [6, 7], rainfall/runoff [8, 9], shallow water condition [10], hydraulic models of flow [11, 12], and further global circulation phenomena [13], including the coupled effects of atmosphere, ocean, and floods [14]. Although physical models showed great capabilities for assessment a diverse range of flooding scenarios, they often require various types of hydro-geomorphological monitoring data sets, requiring intensive computation, which highly complex to model [15]. Furthermore, as stated in Reference [16], the development of physically based models often requires in-depth knowledge and expertise regarding hydrological parameters, reported to be highly challenging. Moreover, numerous studies suggest that there is a gap in assessment capability of physical models [14]. Nevertheless, major improvements in physically based models of flood were recently reported through the hybridization of models [17], as well as advanced flow simulations [18, 19]. In addition to numerical and physical models, data-driven models also have a long tradition in flood modeling, which recently gained more popularity. Data-driven methods of prediction assimilate the measured climate indices and hydro-meteorological parameters to provide better insight. For instance, statistical models of autoregressive moving average (ARMA) [20], multiple linear regression (MLR) [21], and autoregressive integrated moving average (ARIMA) [22] are the most common flood frequency analysis (FFA) methods for modeling flood prediction. FFA was among the early statistical methods for predicting floods [23]. Regional flood frequency analyses (RFFA) [24], more advanced versions, were reported to be more efficient when compared to physical models considering computation cost and generalization. Assuming floods as stochastic processes, they can be predicted using certain probability distributions from historical streamflow data [25]. explained; Hydrological model is an essential tool to help build the linkage between weather information and river runoff. Previously, a wide range of hydrological models have been developed, and most of them describe the rainfall-runoff transformation processes based on physical theory. From the previous studies, it is found that flood's models normally computationally intensive and physically-based hydrological modeling has strict requirement on input data. This poses significant challenges for developing countries where relatively limited resources and data are available. The statistical and/or data-driven tools, which are computationally cheaper and efficient, have gained their popularity in the past decades in the fields of either producing flood reliable map, flood risk assessment or hydrological modeling. The data-driven approaches, like Artificial Neural Network (ANN) and support vector machine (SVM) and decision tree (DT) are found to be effective in improving flood risk assessment especially producing flood reliable map. [26, 27, 28]. In hydrological modeling field, applications of ANN can be found in [29].

According to this document, the flood risk management should be based primarily on the detection of areas prone to flood occurrence. At the same time, the assessment of flood susceptibility is also the most important non-structural measure adopted to reduce the material damage and the loss of human lives generated by these natural hazards.

The use of the latest Geographic Information System technologies allows a fast evaluation of flood susceptibility for a given area. The accuracy of GIS processing is determined in a decisive way by the accuracy of input data, and also by GIS techniques combination with the machine learning and statistical algorithms. Recently, there has been an exponential increase in the number of papers regarding the identification and mapping of flood-prone areas by using the combination between GIS techniques and some models specific to bivariate statistics or machine learning. The Statistical Index, Weights of Evidence and Frequency Ratio are three of the most popular bivariate statistics models used to evaluate the flood susceptibility [30].

The main limitation of the bivariate statistical models is that they take into account only the spatial relationship between the flood locations and conditioning factors, while the relationship between the predictors themselves is not considered [30]. The most commonly used machine learning models for detecting the surfaces prone to flood phenomena are: support vector machine, decision trees algorithms, artificial neural networks [30, 31]. and adaptive neuro-fuzzy inference system. The main advantages of machine learning algorithms reside in their high automation degree, and also, in the easy identification of trends and patterns within a dataset. Also, the machine learning methods can run with multi-variety and with multi-dimensional data. Even if the machine learning algorithms are considered advanced techniques, however, there are also some disadvantages that characterize these models. The large volume of data that should be used to train the models and the high error-susceptibility are two of the drawbacks most discussed in the literature.

Taking all the above mentioned into account, it can be concluded that ML method as powerful method that can has impressive effect to advance flood risk assessment and producing flood reliable map. Hence, in this literature review try to present comprehensive overview in latest advancement of FRA by ML methods. The main structure of this survey are as follows:

## 2. Method and Outline

This survey describes the state of the art of ML methods for flood risk assessment and producing map to identify flood prone areas 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), CiteScore, SCImago journal rank (SJR), and h-index. The papers were reviewed in terms of type of data (conventional and less conventional data as flood resource variables), ML methods, type of ML methods (single and combined methods) and obtained results.

The applications in flood risk assessment can be classified according to type of data, conventional data such as flood inventory map, flood conditions factors (altitude, aspect, slope, curvature, stream power index, topographic wetness index, sediment transport index, topographic roughness index, distance from river, geology, soil, surface runoff, and land use/cover (LULC)), topographic data and rainfall data and less conventional data such as remote sensing data, survey from public, community mapping drainage data and data from social media. 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 [32]. 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 risk assessment may vary significantly according to the application, dataset, and type of ML methods. For instance, ML methods used for flood risk assessment are significantly different from those used prediction. Figure 1 represents the organization of the search queries and further describes the survey search methodology. The search query included three main search terms. The flood resource variables were considered as term 1 of the search (<Flood resource variable1-n>), which included 25 keywords for search queries mentioned above. Term 2 of search (<ML method1-m>) included the ML algorithms. The collection of the references [15,22,24,33,34,35,36] provides a complete list of ML methods, from which the 25 most popular algorithms in engineering applications were used as the keywords of this search. Term 3 included the four search terms most often used in describing flood risk assessment, i.e., "assessment", "management", "forecast", or "analysis". The total search resulted in 483 articles. Among them, 110 original research papers were refined through our quality measure included in the survey.

Section 3 presents the state of the art of ML in flood risk assessment. A technical description on the ML method and a brief background in flood applications are provided. Section 4 Section 5 presents.

## 3. State of the Art of ML Methods in Flood Risk Assessment

For creating the ML assessment, the historical records of flood events, in addition to real-time cumulative data of a number of rain gauges or other sensing devices for various return periods, are often used. The sources of the dataset are traditionally rainfall, Flood inventory map, Topographic data, geological characteristics, soil characteristics, Hydrological data and water level, measured either by ground rain gauges, or relatively new remote-sensing technologies such as satellites, multi sensor systems, and/or radars [37].

Nevertheless, remote sensing is an attractive tool for capturing higher-resolution data in real time. In addition, the high resolution of weather radar observations often provides a more reliable dataset compared to rain gauges [37]. Thus, building assessment model based on a radar rainfall dataset, community mapping drainage data social media-tweeter was reported to provide higher accuracy in general [38]. Whether using a radar-based dataset or ground gauges to create an assessment model, the historical dataset values is divided into individual sets to construct and evaluate the learning models. To do so, the individual sets of data undergo training, validation, verification, and testing. The principle behind the ML modeling workflow and the strategy for flood modeling are described in detail in the literature [39,40]. Figure 2 represents the basic flow for building an ML model. The major ML algorithms applied to flood assessment include Decision Tree (DT) [41, 42, 28], support vector machines (SVM) [41, 43, 27], Artificial Neural Networks (ANNs) [44, 45, 46], Random Forest (RF) [47, 48, 49], Logistic Regression (LR) [50, 51], Frequency Ratio (FR) [52,46] In the following subsections, a brief description and background of these fundamental ML algorithms are presented.

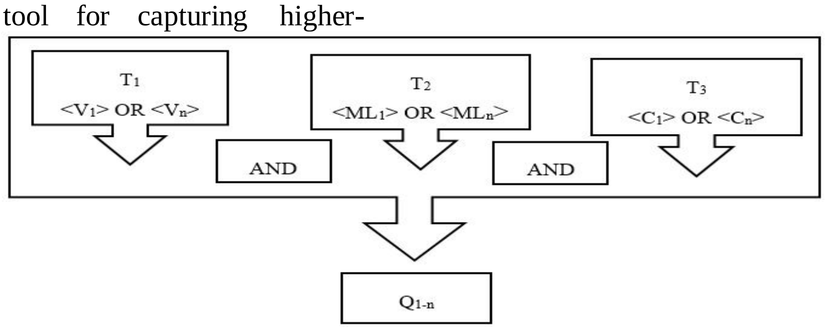


**Figure 1.** Flowchart of the search queries.

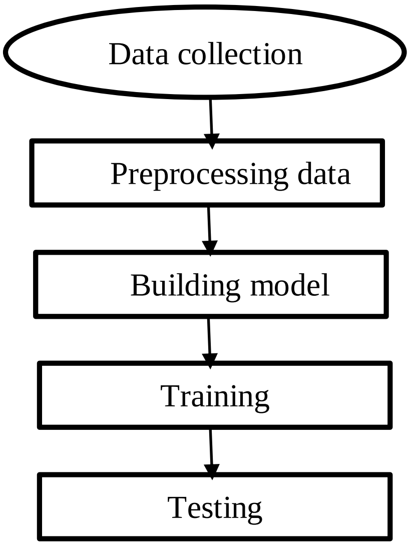


**Figure 2.** Basic flow for building the machine learning (ML) model.

## 4. Machine learning models

The most five applicable machine learning algorithms to assess flood and produce map for identifying flood prone areas are as follows:

### 4.1. Decision Tree (DT)

The ML method of DT is one of the contributors in flood modeling with a wide application in flood simulation. DT is broadly used in classification and modeling. It has hierarchical structure that group the conditioning factors into the homogeneous classes with different susceptibility levels. There is no need for any strict assumptions about the distribution of data in order to perform DT. Furthermore, all the data formats can be used in its analysis such as scale, nominal and etc. DT is selected for comparison with ensemble method as it is one of the robust machine learning methods and its procedure is much easier to be understood compared to other methods like ANN. DT has tree structure which is constructed by a root node, a set of internal nodes, and a set of terminal nodes [50]. The conditioning factors that have significant influence on flooding will be used in the processing, while the others will be rejected by the program. DT can be implemented using various ways such as chi squared automatic interaction detection (CHAID), Exhaustive CHAID, classification and regression trees, Quick, Unbiased, Efficient Statistic Tree and etc. 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) [53]. 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 [26]. The classification and regression tree (CART) [54,55], which is a popular type of DT used in ML, was successfully applied to flood modeling; however, its applicability to flood prediction is yet to be fully investigated [56].

Random forest: A classification and regression technique based on assembling a large number of decision trees. Specifically, it is an ensemble of trees constructed from a training dataset and internally validated to obtain a dependent variable by given independent variables. Two powerful advantages of machine learning techniques are used in RF: bagging and random feature selection. For bagging, each tree is trained on a bootstrap sample of the training data, and predictions are made by a majority vote of trees. Further, the model randomly selects a subspace of feature predictions to split at each node when growing a tree [57]. It is known as popular DT method for flood modeling [47]. RF includes a number of tree predictors. Each individual 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 [58]. Reference [59] introduced RF as an effective alternative to SVM, which often delivers higher performance in flood prediction modeling. [60] compared the performances of ANN, SVM, and RF in general applications to floods, whereby RF delivered the best performance. Another major DT is the M5 decision-tree algorithm [61]. M5 constructs a DT by splitting the decision space and single attributes, thereby decreasing the variance of the final variable. Further DT algorithms popular in flood risk assessment include reduced-error pruning trees (REPTs), Naïve Bayes trees (NBTs), chi-squared automatic interaction detectors (CHAIDs), logistic model trees (LMTs), alternating decision trees (ADTs), and exhaustive CHAIDs (E-CHAIDs). Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

**Put simply:** random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Let's look at random forest in classification, since classification is sometimes considered the building block of machine learning. Random forest has nearly the same hyper parameters as a decision tree or a bagging classifier. Fortunately, there's no need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest. With random forest, you can also deal with regression tasks by using the algorithm's regression.

Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Therefore, in random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).

Boosted regression tree is another popular DT method for flood modeling (belonging to the gradient boosting modeling family): A tree-based model that combines a large number of machine learning and regression tree models to learn and weigh them (by assigning individual weights to every sample point of the training dataset), in order to describe the relationship between the independent and dependent variables. It uses several techniques to improve the performance of a single model, e.g., by creating an ensemble of regression models [62].

### 4.2. Support vector machine (SVM)

SVM is a binary classifier and machine learning algorithm which is a new supervised learning method and is based on the structural risk minimization principle [26]. Separating hyper-plane creation from training dataset is the basis of this method [26]. SVMs are popular because of good empirical performance (compared with other models, such as artificial neural networks), easy training process, and the avoidance of local minima, relatively suitable mathematics for multi-dimensional data, and a tradeoff between complexity and error. [63] proposed and classified the support vector

(SV) as a nonlinear search algorithm using statistical learning theory. Later, the SVM [64] was introduced as a class of SV, used to minimize over-fitting and reduce the expected error of learning machines. SVM is greatly popular in flood modeling; it is a supervised learning machine 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 on training from past data. Over the past two decades, the SVM was also extended as a regression tool, known as support vector regression (SVR) [65].

SVMs are today know as robust and efficient ML algorithms for flood risk assessment and producing flood reliable map [27]. SVM and SVR emerged as alternative ML methods to ANNs, with high popularity among hydrologists for flood risk assessment. They use the statistical learning theory of structural risk minimization (SRM), which provides a unique architecture for delivering great generalization and superior efficiency. Most importantly, SVMs are both suitable for linear and nonlinear classification, and the efficient mapping of inputs into feature spaces [66]. Thus, they were applied in numerous flood assessment cases with promising results, excellent generalization ability, and better performance, compared to ANNs and MLRs, e.g., extreme rainfall [67], precipitation [68], rainfall–runoff [69], reservoir inflow [70], streamflow [71], flood quantiles [39], flood time series [72], and soil moisture [73]. Unlike ANNs, SVMs are more suitable for nonlinear regression problems, to identify the global optimal solution in flood models [74].

Although the high computation cost of using SVMs 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 [75]. The alternative approach of LS-SVM involves solving a set of linear tasks instead of complex quadratic problems [76]. Nevertheless, there are still a number of drawbacks that exist, especially in the application of flood risk assessment, producing reliable flood map and prediction of flood using LS-SVM [77].

### 4.3 Artificial Neural Networks (ANNs)

ANNs are efficient mathematical modeling systems with efficient parallel processing, enabling them to mimic the biological neural network using interconnected neuron units. Among all ML methods, ANNs are the most popular learning algorithms, known to be versatile and efficient in modeling complex flood processes with a high fault tolerance and accurate approximation [78]. In comparison to traditional statistical models, the ANN approach was used for prediction with greater accuracy [79]. ANN algorithms are the most popular for modeling flood assessment since their first usage in the 1990s [80]. Instead of a catchment's physical characteristics, ANNs derive meaning from historical data. Thus, ANNs are considered as reliable data-driven tools for constructing black-box models of complex and nonlinear relationships of rainfall and flood [81], as well as river flow and discharge forecasting [82]. Furthermore, a number of surveys (e.g., Reference [83]). Suggest ANN as one of the most suitable modeling techniques which provide an acceptable generalization ability and speed compared to most conventional models. References [84,85] provided reviews on ANN applications in flood. ANNs were already successfully used for numerous flood prediction applications, e.g., streamflow forecasting [86], river flow [87,88], rainfall–runoff [89], precipitation–runoff modeling [90], water quality [91], evaporation [92], river stage prediction [93], low-flow estimation [94], river flows [95], and river time series [86]. Despite the advantages of ANNs, there are a number drawbacks associated with using ANNs in flood modeling, e.g., network architecture, data handling, and physical interpretation of the modeled system. A major drawback when using ANNs is the relatively low accuracy, the urge to iterate parameter tuning, and the slow response to gradient-based learning processes [96]. Further drawbacks associated with ANNs include precipitation prediction [97,98] and peak-value prediction [99].

The feed-forward neural network (FFNN) [21] is a class of ANN, whereby the network's connections are not in cyclical form. FFNNs are the simplest type of ANN, whereby information moves in a forward direction from input nodes to the hidden layer and later to output nodes. On the other hand, a recurrent neural network (RNN) [100] is a class of ANN, whereby the network's connections form a time sequence for dynamic temporal behavior.

Furthermore, RNNs benefit from extra memory to analyze input sequences. In ANNs, backpropagation (BP) is a multi-layered NN where weights are calculated using the propagation of the backward error gradient. In BP, there are more phases in the learning cycle, using a function for activation to send signals to the other nodes. Among various ANNs, the backpropagation ANN (BPNN) was identified as the most powerful prediction tool suitable for flood time-series prediction [22]. Extreme learning machine (ELM) [101] is an easy-to-use form of FFNN, with a single hidden layer. Here, ELM was studied under the scope of ANN methods. ELM for flood prediction recently became of interest for hydrologists and was used to model short-term streamflow with promising results [102,103].

### 4.4. Logistic Regression (LR)

LR is a multivariate statistical model for flood susceptibility mapping [104]. The benefit of this model is that data do not require to be normally distributed and the factors can either be categorical, continuous, or any combination of both [104]. In this model, flood susceptibility map developed from flood inundation area was considered as the dependent variable, where 1 is for flooding area and 0 is for non-flooding area. The mathematical expression of the LR model is given by: [30, 105]

$$P = \frac{1}{1+e^{-z}} = \frac{1}{\left[1+e\left\{-\left(b_0+b_1x_1+....+b_nx_n\right)\right\}\right]} \quad (1)$$

where P is the probability of occurrence of flood or non-flood, z is the linear combination, n is the number of flood conditioning factors, $x_i\left(i=1,..,n\right)$ is the flood conditioning factors, b0 is the intercept of the model, $b_i\left(i=0,1,..,n\right)$ is the regression coefficients for the independent variables of the logistic regression model

### 4.5. Frequency Rate (FR)

FR model is based on the observed relationships between the distribution of the floods and flood conditioning factor [104]. The frequency ratio for the class of each conditioning factor was calculated by dividing the flood occurrence ratio by the area ratio. Each factor frequency ratio was calculated using Eq. (2) and flood susceptibility map was developed from Eq. (3) [104].

$$FR = \frac{percentage\ of\ flood}{percentage\ of\ the\ class\ of\ each\ conditioning\ factor} \quad (2)$$

$$FS = FR_1 + FR_2 + .... + FR_n \quad (3)$$

Where FR is the ranking of each conditioning factors, n is the number of total factors for flood susceptibility (FS). When assessing the probability of flooding during the particular time period and in a specific environment, it is essential to distinguish the conditioning factors and the situations that can cause flooding. In order to perform the flood susceptibility analysis, the FR method was applied using GIS techniques. FR is very robust method for the purpose of BSA, as it considers the impact of each conditioning factor on flooding and assigns the weights very precisely. FR method is based on the relationship between spreading of flooding and each conditioning factor, to exhibit the relationship between flood locations and the conditioning factors in the study area. If the value of FR be more than 1, it means the percentage of the flooding is greater than the area and shows greater correlation; however, values less than 1 represent a minor correlation.FR ratios for conditioning factors were calculated and the weights were normalized in the range 0–1. The normalization should be done because the ten conditioning factors vary in dimensions and are not appropriate for direct input for the SVM model. For normalization process, a common method was used for this purpose.

For example, for $y_i \left(i = 1,..,n\right)$

$$Y_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \quad (4)$$

Where, $y_i$ means the normalized values of $Y_i$, $y_{min}$ and $y_{max}$ represent the minimum and maximum value of $Y_i$ respectively. In this manner, the nominal and interval class group data were converted to scale values ranging from 0 to 1. Therefore, these scale conditioning factors were normalized and classified based on the weights of FR and then feed into SVM model.

## 5. Classification of ML Methods and Applications

The most popular ML modeling methods for flood risk assessment were identified in the previous section, including DT, SVM, ANNs, RF and LR. Figure1 presents the major ML methods used for flood risk assessment, and the number of corresponding articles in the literature over the last decade. This figure was designed to communicate to the readers which ML methods increased in popularity among hydrologists for flood modeling within the past decade.

Considering the ML methods for application to floods, it is apparent that DT, SVM, ANNs, RF and LR are the most popular. These ML methods can be categorized as single and hybrid methods. In addition to the fundamental hybrid ML methods, i.e., DT, SVM, ANNs, several different research strategies for obtaining better assessment evolved [46,47]. The strategies involved developing hybrid ML models using soft computing techniques, statistical methods, and physical models rather than individual ML approaches, whereby the extra components complement each other with respect to their drawbacks and shortcomings. The success of such hybrid approaches motivated the research community to explore more advanced hybrid models. Figure 4 presents the progress of single and hybrid ML methods for flood risk assessment in the literature over the past decade. The figure shows an apparent continuous increase and notable progress in using novel hybrid methods.

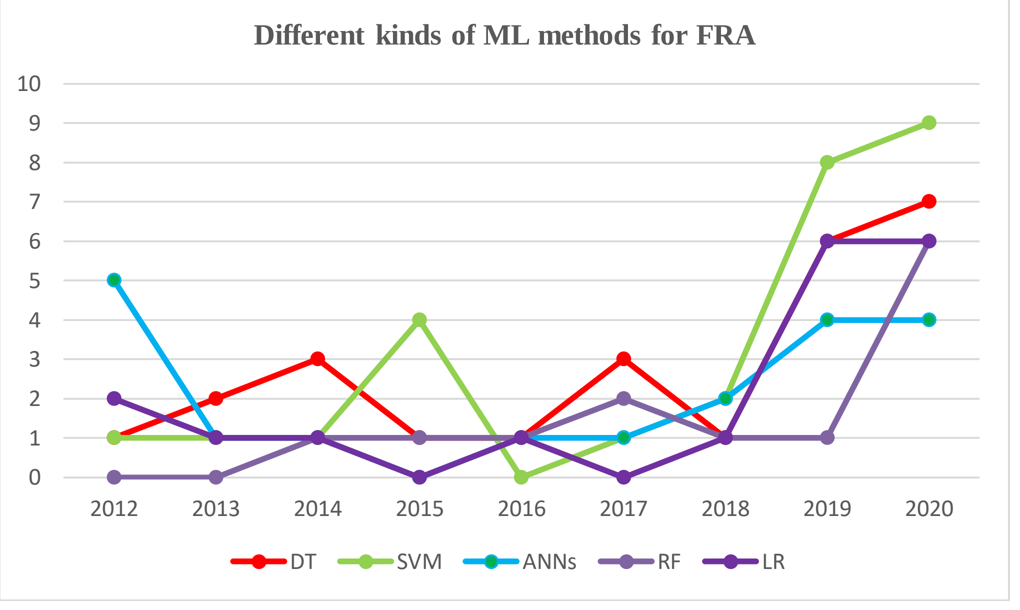


**Figure3:** The most popular model for FRA in the literature review from 2012-May 2020

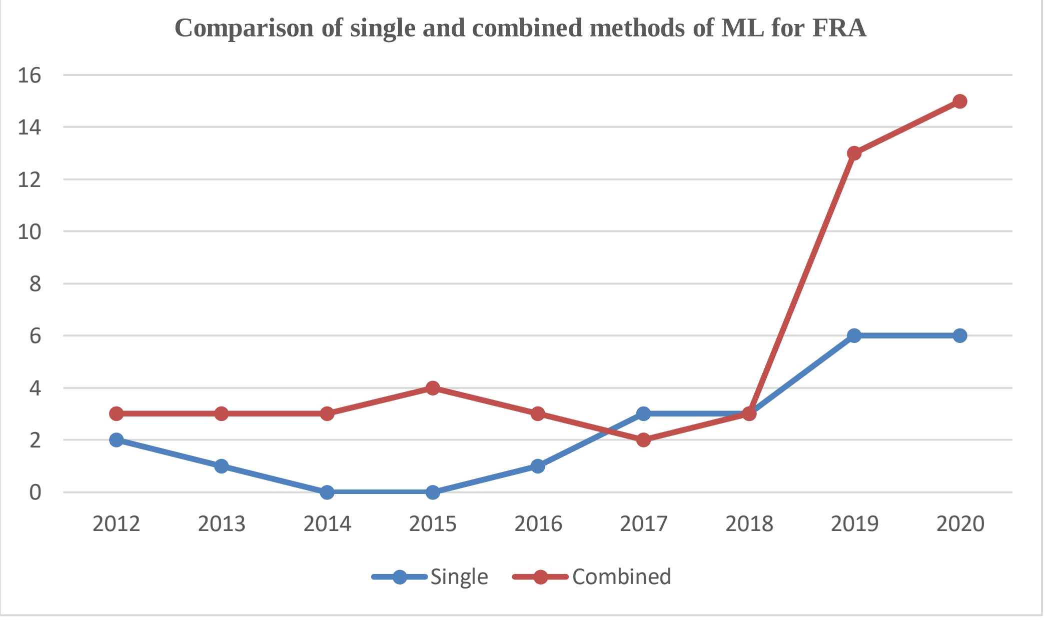


**Figure 4:** The progress of single and combined ML methods for FRA in the literature review from 2012-May 2020

## 6. Type of Data:

The data that is used for flood risk assessment just was conventional data but in recent decade simultaneous with promotion of technology and new device that can impressive effect to increase the accuracy of measurement for specific purpose the less conventional data such as participation of public to gather information ,community mapping drainage data, data from social media and data from satellite, radar or remote sensing especially in areas that suffer from data scarcity has had eye catching progress to assess flood or producing reliable flood map.

The data which is used in this literature review is divided three sections based on resource:

### 6.1. Conventional data

The traditional data that is used to flood modeling such as flood inventory map, flood resource variable (altitude, slope, aspect, curvature, distance from river, topographic wetness index (TWI), drainage density, soil depth, soil hydrological groups (SHG), land use and lithology), Rain fall data, water level and so on. In below some of them are described.

#### 6.1.1. Flood Inventory Map

To evaluate flood risk in an area, analyzing records of past flood events is essential. Therefore, an inventory map is considered the most essential factor for predicting future disaster occurrence; such map can represent single or multiple events in a specific area. Thus, the first stage in flood susceptibility analysis is to acquire information about the floods that have occurred in the past. Since flood occurrences in the past and present are keys to future spatial prediction, a flood inventory map is a prerequisite for such a study. Accurate detection of the flood locations is so important for probabilistic flood hazard analysis [1, 106]. The flood inventory map is a basic map for flood susceptibility assessment [30]. Accurate analysis of flood susceptibility requires a precise flood inventory map that shows the locations of flood occurrence. [52]

#### 6.1.2. Flood resource variables

The data available in the literature review is dependent on case study areas and usually the effective flood susceptibility parameters were identified: altitude, slope, aspect, curvature, distance from river, topographic wetness index (TWI), drainage density, soil depth, soil hydrological groups (SHG), land use, lithology and rainfall data. In next section, all probabilistic flood condition factors are explained briefly. [106]

**Altitude:** Altitude is one of the most important factors affecting flooding [30]. In general, there is an inverse relationship between flood and elevation; flood frequency increases with decreasing elevation, with the result that lower elevations are more susceptible to flooding.

**Slope:** Flooding is directly related to slope gradient and is an important physiographic characteristic [60]. Slope contributes directly to surface runoff velocity and vertical percolation and thus affects flood susceptibility.

**Aspect:** The slope aspect is defined as the direction of the maxi-mum slope of the terrain surface, which in most flood studies was considered an important flood susceptibility parameter.

**Curvature:** Curvature was determined to be another influential conditioning parameter and was extracted from the DEM map in ArcGIS. The curvature consists of three categories: concave, convex, and flat surface. It is a factor in run off flow and can be useful in detecting flood susceptibility.

**Distance from river:** River flows are the main pathways for flood discharge and areas near rivers are susceptible to flooding.

**Topographic wetness index (TWI):** This parameter was presented by and indicates the spatial variations of wetness in a watershed. In other words, this index exhibits the amount of water accumulation in any pixel size of the watershed area, which is calculated as:

$$TWI - In\left(\frac{A_s}{\tan\beta}\right) \quad (5)$$

Where $A_s$and β are the specific catchment area ($m^2m^{-1}$) and the slope gradient (in degrees), respectively. TWI was calculated in the SAGA GIS environment.

**Drainage density:** When rainfall occurs in a watershed, the drainage density has effects on peak flows. Regional floods are often related to peak discharge value and drainage area; therefore, this parameter has an important influence on flood susceptibility. Poor drainage systems often result in river overflow and continuous flooding in an area. The drainage density was derived from the line density tool using ArcGIS software.

**Soil type:** Soil depth, soil texture, and soil porosity are the main factors affecting surface runoff. They mainly affect runoff generation by changing the infiltration characteristics and water holding characteristics of the soil. Different soil types indicate different soil properties, so soil type was selected as a conditioning factor. [107].

**Soil depth:** Soil depth was considered as the depth of the soil layer from the ground surface to bedrock. In some areas with low soil depth and especially upstream, runoff generation is higher. [106]

**Soil hydrological groups:** (SHG). Soil hydrological groups indicate soil quality that is based on a minimum water infiltration rate. Soil hydrological groups are classified into four groups: A, B, C,and D. Soils in group A have the minimum runoff potential, while soils in group D have the maximum runoff potential.

**Land use:** Land use has a significant role in runoff speed, interception, infiltration, and evapo-transportation. The land use map is essential for the determination of areas susceptible to flooding.

**Lithology:** Lithology has a significant influence on hydrological processes in a watershed. Different lithology units have different susceptibilities to flooding.

**Rainfall data:** Rainfall was selected as an essential factor because floods are often triggered by high-intensity and short rainstorms [27].

### 6.2. Less conventional data

Many parts of the world are, however, still lacking basic geographical and hydrological data, given the traditionally high barriers to produce it. Furthermore, the urbanization is so rapid in many areas that it is non-beneficial to carry out expensive surveys that soon get outdated [6]. In the last decade, the development of Internet and smartphones have made public participation in monitoring, data collection, planning and decision making possible in a way that has never been accessible previously. Paul et al. [78] show how a polycentric approach is beneficial over a monocentric, both in pre-disaster, in-disaster and post-disaster management. They argue that a participatory approach to data collection can support multidirectional information provision and enhance hydrological risk reduction [1]. state that most of the data needed for flood risk management is found at the local scale, and that the lead time of top-down implemented environmental policies are usually too long to decrease the vulnerability of people living in flood prone areas. One way of involving citizen in data collection for building disaster resilience is the emerging field of community mapping. Data from satellite or remote sensing and data from social media- tweeter are the other type of less conventional data that will be described in the next section.

### 6.2.1. Volunteered Geographic Information

Community mapping is the action of producing a map together with or by the residents of a certain location, often featuring local knowledge and resources [35]. It can be understood as a form of Volunteered Geographic Information (VGI), a term coined by Goodchild in 2007 [36]. Being the very first notion of this emerging field, it is worth quoting here; they [the mapping volunteers] are largely untrained and their actions are almost always voluntary, and the results may or may not be accurate. But collectively, they represent a dramatic innovation that will certainly have profound impacts on geographic information systems (GIS) and more generally on the discipline of geography and its relationship to the general public. I term this volunteered geographic information (VGI), a special case of the more general Web phenomenon of user-generated content [99]. VGI, in turn, can be seen as crowdsourcing of geospatial information. Felstiner [91] defines crowdsourcing as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." Examples of crowdsourced data are Wikipedia, the free online encyclopedia, and Open StreetMap3 (OSM), an online world map which can be edited by anyone [2]. The field of VGI, as defined by Goodchild, has emerged as a consequence of Web 2.0, that is the development of the user-generated World Wide Web [68, 108]. Since the early 2000's, it is increasingly available for anyone with Internet access to create geographic information, which was accommodated by several coincides around the new millennium. The Global Positioning System (GPS) was fully operational in 1995, implemented by the US Department of Defense, and the selective availability of the GPS signal was removed in the year 2000. This gave way to affordable GPS receivers, and home computers and user-generated websites were on the rise [6]. VGI has put mapping, a task that for centuries has been reserved to official agencies, in the hands of anyone that want to contribute to online maps [109]. This section will introduce different forms of VGI, namely geographical citizen science, crisis mapping and community mapping.

Volunteered data can be diverse, complex, and overwhelming in volume, velocity, and in the variety of viewpoints they offer [29]. Negotiating these overwhelming streams is beyond the capacity of human analysts. Current research offers some novel capabilities to utilize these streams in new, groundbreaking ways, leveraging, fusing, and filtering this new generation of air-, space-, and ground-based sensor-generated data. This research presents a novel approach to prioritizing the collection of remote sensing data from satellites, airplanes, and UAVs during hazard events by utilizing VGI as a filtering tool. In addition, it proposes the use of VGI for disaster assessment to fill in the gaps when remote-sensing data are lacking or incomplete. In order to use social media efficiently and effectively and VGI to 'cue' or augment satellite observations, it is necessary to filter the data for content and to geolocation them using a variety of text mining and network analysis algorithms. Filtering yields a rapid and direct identification of affected areas, which can aid authorities to prioritize site visits and response initiatives as well as to task additional data collection.

### 6.2.2. Participation of public in data collection

The participation of the general public in the research design, data collection and interpretation process together with scientists is often referred to as citizen science. While citizen science itself has existed since the start of scientific practice, developments in sensing technology, data processing and visualization, and communication of ideas and results, are creating a wide range of new opportunities for public participation in scientific research. In the literature reviews the state of citizen science in a hydrological context and explores the potential of citizen science to complement more traditional ways of scientific data collection and knowledge generation for hydrological sciences and water resources management. Although hydrological data collection often involves advanced technology, the advent of robust, cheap, and low-maintenance sensing equipment provides unprecedented opportunities for data collection in a citizen science context. These data have a significant potential to create new hydrological knowledge, especially in relation to the characterization of process heterogeneity, remote regions, and human impacts on the water cycle. However, the nature and quality of data collected in citizen science experiments is potentially very different from those of traditional monitoring networks. This poses challenges in terms of their processing, interpretation, and use, especially with regard to assimilation of traditional knowledge, the quantification of uncertainties, and their role in decision support. It also requires care in designing citizen science projects such that the generated data complement optimally other available knowledge.

### 6.2.3. Social media

Novel information streams, such as social media-contributed videos, photographs, and text as well as other open sources, are redefining situation awareness during emergencies. When these contributed data contain spatial and temporal information, they can provide valuable Volunteered Geographical Information (VGI), harnessing the power of 'citizens as sensors' to provide a multitude of on-the-ground data, often in real time. There are several opportunities and challenges associated with the use of VGI. Examine the content and characteristics of VGI, the technical and social processes through which it is produced, appropriate methods for synthesizing and the use of these data in research, and emerging social and political concerns related to this new form of information. Although these volunteered data are often published without scientific intent, and usually carry little scientific merit, it is still possible to mine mission critical information. For example, during hurricane Katrina, geolocated pictures and videos searchable through Google provided early emergency responders with ground view information. These data have been used during major events, with the capture, in near-real-time, of the evolution and impact of major hazards. Specifically, VGI based on Twitter and other non-authoritative data have been shown to contain valuable data that can be used for improving flood estimation in near-real-time. Volunteered data can be employed to provide timely damage assessment, help in rescue and relief operations, as well as for the optimization of engineering reconnaissance While the quantity and real-time availability of VGI make it a valuable resource for disaster management applications, data volume, as well as its unstructured, heterogeneous nature, makes the effective use of VGI challenging.

Recently, in Europe and US the use of social science research approaches and methods has a significant contribution in flash flood risk reduction and warning. Engineers and physical scientist now recognized that their elucidations cannot be operated without engaging social science. The S-P-R-C model is favorable to understand the concept of hazard, vulnerability and risk. In assessing the risk, there must be hazard and vulnerability (source/initiator, pathway and receptors). The consequences depend on the exposure of the receptors to the hazard. Historical flash flood events are shown that deaths and property losses highly significant which are because of ineffective flash flood risk assessment approaches. There are four important steps in risk assessment; first step is based on characterization of area such as physical, social, geomorphological, hydro-meteorological, land use land cover and analysis of historical events. Secondly, determining the intensity of a flash flood (the strength of the flash flood) and developing alternative scenarios in the basin. Thirdly, vulnerability assessment which is based on physical vulnerability (susceptibility and exposure) and socioeconomic vulnerability (qualitative and quantitative indicators). Fourth, risk assessment is the combination of hazard intensity level, scenario, and total physical and socioeconomic vulnerability. Flash flood risk assessment is critical to management and planning for the future changes in frequency and magnitude such events.

Twitter is one of the largest social networking sites, and it is widely used to share information through micro-blogging. These micro-blogs, or ‘tweets’, are limited to 140 characters, so abbreviations and colloquial phrasing are common, making the automation of filtering by content challenging. Twitter is very popular during emergencies and disasters, and it is being used by both official government agencies and the public to disseminate information. Central to the operation of Twitter is the use of hashtags, words or un spaced phrases prefixed with the sign #. They are identifiers unique to Twitter and are frequently used to search and filter information. The creation and use of a hashtag can be established by any user who wants to create a concept category to share specific information about a subject. For example, during the 2013 Boulder floods, the hashtag #boulder flood was used by users and agencies to share information about this particular event. Twitter data can be queried for specific hashtags or text present in the tweets, and for spatial and temporal constraints. There are several web-based tools and an API for the automatic querying, filtering, and displaying of tweets. For more literature review, tweets are harvested using the CarbonScanner application to identify ‘hotspots’ and task satellite data collection. CarbonScanner scans tweets, identifies relevant keywords and hashtags, and georectifies the data.

### 6.2.4. Remote Sensing

Every year natural hazards are responsible for powerful and extensive damage to people, property, and the environment. Drastic population growth, especially along coastal areas or in developing countries, has increased the risk posed by natural hazards to large, vulnerable populations at unprecedented levels. Furthermore, unusually strong and frequent weather events are occurring worldwide, causing floods, landslides, and droughts affecting thousands of people. A single catastrophic event can claim thousands of lives, cause billions of dollars of damage, trigger a global economic depression, destroy natural landmarks, render a large territory uninhabitable, and destabilize the military and political balance in a region. Furthermore, the increasing urbanization of human society, including the emergence of megacities, has led to highly interdependent and vulnerable social infrastructure that may lack the resilience of a more agrarian, traditional society. In urban areas, it is crucial to develop new ways of assessing damage in real-time to aid in mitigating the risks posed by hazards. Annually, the identification, assessment, and repair of damage caused by hazards requires thousands of work hours and billions of dollars.

Remote sensing data are of paramount importance during disasters and have become the de-facto standard for providing high resolution imagery for damage assessment and the coordination of disaster relief operations. First responders rely heavily on remotely sensed imagery for coordination of relief and response efforts as well as the prioritizing of resource allocation. Determining the location and severity of damage to transportation infrastructure is particularly critical for establishing evacuation and supply routes as well as repair and maintenance agendas. Following the Colorado floods of September 2013 over 1000 bridges required inspection and approximately 200 miles of highway and 50 bridges were destroyed.1 A variety of assessment techniques were utilized following Hurricane Katrina in 2005 to evaluate transportation infrastructure including visual, non-destructive, and remote sensing. However, the assessment of transportation infrastructure over such a large area could have been accelerated through the use of high resolution imagery and geospatial analysis. Despite the wide availability of large remote sensing datasets from numerous sensors, specific data might not be collected in the time and space most urgently required. Geo-temporal gaps result due to satellite revisit time limitations, atmospheric opacity, or other obstructions. However, aerial platforms, especially Unmanned Aerial Vehicles (UAVs), can be quickly deployed to collect data about specific regions and be used to complement satellite imagery. UAVs are capable of providing high resolution, near real-time imagery often with less expense than manned aerial- or space-borne platforms. Their quick response times, high maneuverability and resolution make them important tools for disaster assessment. Contributed data that contain spatial and temporal information can provide valuable Volunteered Geographic Information (VGI), harnessing the power of ‘citizens as sensors’ to provide a multitude of on-the-ground data, often in real time. Although these volunteered data are often published without scientific intent, and usually carry little scientific merit, it is still possible to mine mission critical information. For example, during hurricane Katrina, geolocated pictures and videos searchable through Google provided early emergency response with ground-view information. These data have been used during major events, with the capture, in near real-time, of the evolution and impact of major hazards.Volunteered data can be employed to provide timely damage assessment, help in rescue and relief operations, as well as the optimization of engineering reconnaissance. While the quantity and real-time availability of VGI make it a valuable resource for disaster management applications, data volume, as well as its unstructured, heterogeneous nature, make the effective use of VGI challenging. Volunteered data can be diverse, complex, and overwhelming in volume, velocity, and in the variety of viewpoints they offer. Negotiating these overwhelming streams is beyond the capacity of human analysts. Current researchs offers some novel capabilities to utilize these streams in new, groundbreaking ways, leveraging, fusing and filtering this new generation of air-, space and ground-based sensor-generated data**.**

Flash floods frequently occur in small catchments or in a small dry land, such areas often poorly gauged or ungauged. Hence, quality of remote sensed data is critical to flash flood forecasting. The main causes are thunderstorms, monsoon trough, rapid melting of snow and glacier lake outburst flooding. Increasing population growth and climate change impacts will be increased risk of more frequent and severe flash floods in the future. Therefore, flash flood hazard vulnerability and risk analysis need special attention in order to reduce severe losses.

### 6.3. Data fusion (Combination of conventional and less conventional data)

White [14] defined data fusion in the book "Data Fusion Lexicon" as "a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results.". [1] thought that "information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making." Data fusion for flood risk assessment is combining conventional and less conventional in order to increase the accuracy for flood risk assessment and producing flood reliable map. In this section the introduction of conventional data and less conventional data are explained, also The figure 5 that is presented in this section can show the type of data that is used in recent decade and how it can be effective on accuracy of flood risk assessment and producing flood reliable map.

## 7. The potential of most popular ML Method for flood risk assessment and producing flood reliable map

This section can introduce perfectly some of research that lead to significant results in flood modeling. Limited research papers are done this field. In below some tables are shown that describe the art of most popular ML methods to

improve flood risk modeling and producing reliable flood map in recent decade.

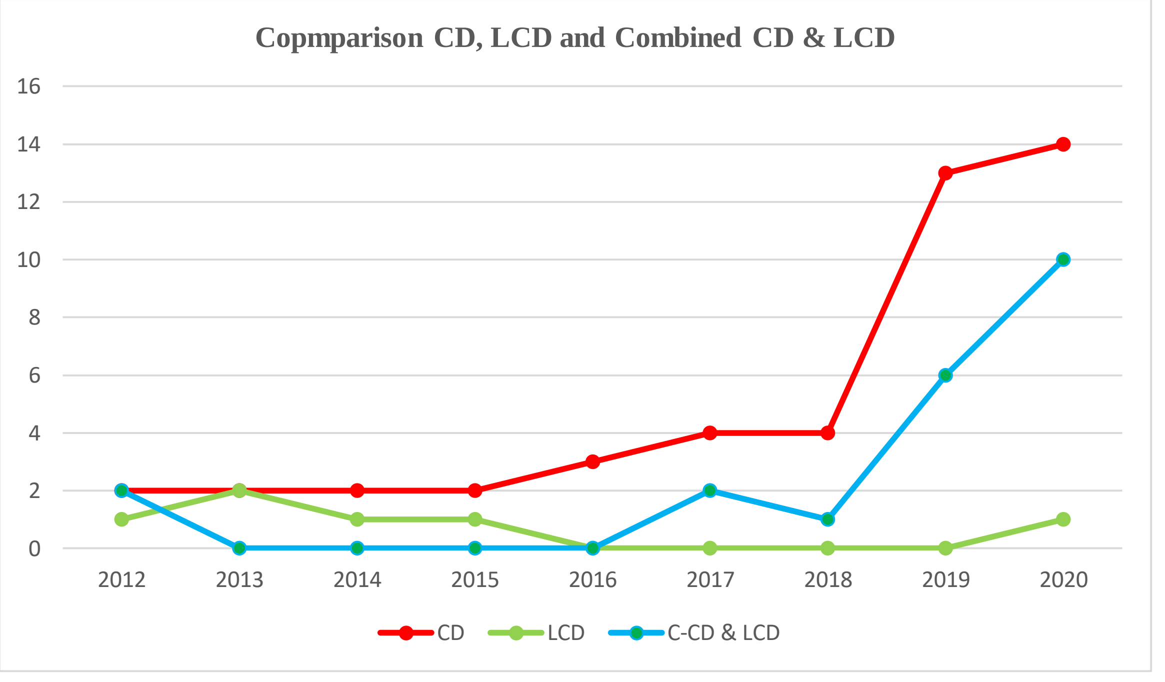


**Figure 5.** Comparison of progress of using conventional-less conventional data and fusion data in recent decade

**1.** The potential of DT algorithm for FRA/FRM and producing map to identify flood prone area.

| REFERENCE | ADVANTAGES |
|---|---|
| **1-JUNFEI CHEN, 2012** | 1-Verification the model's (Diffusion Tress (DT)) effectiveness for flood risk management.<br>2-This method can be easily applied to effectively resolve problems of insufficient samples in flood risk assessment. |
| **2-MAHYAT SHAFAPOUR TEHRANY, 2013** | 1-The validation results showed that, area under the curve for the results of DT and integrated method of FR and LR was 87% and 90% for success rate and 82% and 83% for prediction rate respectively.<br>2-DT is based on the rules which are created precisely and strongly by considering all the characteristics of the variables which can enhance the performance of the flood susceptibility mapping. |
| **3-PETER LAMOVEC, 2013** | The best results are produced by the J48 decision tree algorithm to reduce the time necessary for flood mapping. |
| **4-MICHELLE WOODWARD,2014** | 1-DT methodology that has the capability to assess the most appropriate set of interventions to make in a flood system,<br>2-The opportune time to make these interventions, given the future uncertainties. |
| **5-KAMRAN CHAPI,2017** | 1-Bagging-LMT (Combination of bagging ensemble and Logistic Model Tree (LMT)) outperformed all these models in terms high prediction accuracy and high quality of flood susceptibility map.<br>2-The Proposed Bagging- LMT model is a better alternative for flood susceptibility modeling.<br>3-The new proposed model is recommended as a suitable method for hydrological studies and disaster management to produce precise and reliable susceptibility map can decrease the cost and damage from environmental disasters, such as floods. |
| **6-KHABAT KHOSRAVI, 2018,** | 1- Advanced DT methods are promising methods for flood susceptibility assessment<br>2- The Alternating Decision Trees (ADT) model has the highest prediction capability for flash flood susceptibility assessment, followed by the Naïve Bayes Trees (NBT), the Logistic Model Trees (LMT), and the Reduced Error Pruning Trees (REPT) respectively.<br>3- ADT, NBT, and LMT machine learning methods are better than conventional methods (FR, W-of-E, and AHP). |

| | |
|---|---|
| **7-SAEID JANIZADEH, 2019** | 1-Based on a variety of performance metrics, ADT method was dominant over the other methods (FT, KLR, MLp and QDA) for flash flood susceptibility map.<br>2-The Functional Tree (FT), method was ranked as the second-best method, followed by the KLR, MLP, and QDA |
| **8-MASAHIKO HARAGUCHI, 2019** | Decision Tree (DT) model is applied to illustrate how developing countries with limited datasets and capacity can utilize global hazard data to support risk-informed decision-making at the local level. |
| **9-MAHYAT SHAFAPOUR TEHRANY,2019** | 1- DT produced the highest accuracies of prediction the final susceptibility mapping result, respectively, using DS1 (the LiDAR dataset). |
| **10-BAHRAM CHOUBIN, 2019** | 1-The ensemble modeling approach (Multivariate Discriminant analysis (MDA), Classification and Regression Trees (CART), and Support Vector Machine (SVM)), As Ensemble modeling approach indicated the novel method are so powerful to identifying flood prone area.<br>2-It can help the manager that these areas should, therefore, be prioritized for the prevention and remediation of floods |
| **11-KHABAT KHOSRAVI, 2019** | Among the models, Naïve Bayesian Tree (NBT) model is a promising tool for the assessment of flood-prone areas and can allow for proper planning and management of flood hazards. |
| **12-WEI CHENA, 2019,** | Comparison with three models: Reduced Error Pruning Trees (REPTree) with Bagging (Bag-REPTree) and Random Subspace (RS-REPTree)<br>1-Results show that the RS-REPTree model has the highest prediction capability for flood susceptibility assessment, for the training and validation datasets, it was followed by the Bag-REPTree and REPTree models, respectively.<br>2-The results also proved the superiority of the ensemble method over using these methods individually. |
| **13-BINH THAI PHAM, 2020** | 1-Performance of Logistic Model Tree (LMT) algorithm is the best among the other four applied methods in this research paper.<br>2-Useful for the construction of accurate flash flood susceptibility maps with the objectives of identifying flood-susceptible areas/zones for proper flash flood risk management |
| **14-BAHRAM CHOUBIN, 2020** | The Bagged Classification and Regression Trees (Bagged CART), and RF models have the same performance and higher than the Mixture Discriminate Analysis (MDA) model for the hazard prediction including flood |
| **15-ROMULUS COSTACHE, 2020** | 1-Two ensemble models, the Alternating Decision Tree (ADT) -Index of Entropy (IOE), (ADT-IOE) and the ADT-AHP have high prediction performance and outperform the other models.<br>2-The ADT-IOE and ADT-AHP are new and promising tools for flash-flood susceptibility modeling. |
| **16-KE XIONG, 2020** | Boosted regression Tree (BRT) model outperformed other models and was the ideal choice for debris flow susceptibility mapping. |

2.The potential of SVM algorithm for FRA/FRM and producing map to identify flood prone area.

| **REFERENCE** | **ADVANTAGES** |
|---|---|
| **1-Z. VOJINOVIC, 2013** | 1-SVM modelling technique has proven to be capable to reconstruct geometry of nonlinear datasets which clearly demonstrates their versatility in the bathymetry estimation applications.<br>2-Development a bathymetry dataset that by SVM model can be used to set up physically-based models for coastal flood modelling work |
| **2-MAHYAT SHAFAPOUR TEHRANY, 2014** | Through this integration, the weak point of WoE can be solved and the performance of the SVM will be enhanced for Flood susceptibility mapping.<br>2-The validation results proved the strength and efficiency of the ensemble method (WoE and RBF-SVM) over the individual methods.<br>3- Success rate and prediction rate for ensemble WoE and RBF-SVM method were 96.48% and 95.67% respectively.<br>5-The proposed ensemble flood susceptibility mapping method could assist researchers and local governments in flood mitigation strategies. |
| **3-GARETH IRELAND, 2015,** | Successful application of SVMs for classifying flooded areas in Landsat imagery to produce accurate map. |
| **4-MAHYAT SHAFAPOUR TEHRANY, 2015** | 1-SVM technique is an efficient and reliable tool in flood susceptibility assessment.<br>2-The resultant flood susceptibility maps can be beneficial in flood mitigation strategies. |
| **5-ZHAOLI WANG, 2015** | 1-Identification of top five most important causative factors of flood out of the eleven risk indices,<br>2-Provide a novel and highly successful approach to flood hazard risk assessment and management for prevention, and reduction of natural disasters in the study basin. |
| **6-MAHYAT SHAFAPOUR TEHRANY, 2015** | Efficiency of the proposed ensemble method (SVM-FR) as rapid, accurate and reasonable in flood susceptibility assessment. |

| | |
|---|---|
| **7-HOSSEIN MOJADDADI,2017** | 2-The ensemble method of FR and SVM can be efficiently used in flood hazard studies because of its simple structure and robust performance for flood risk assessment. |
| **8-VIET-NGHIA NGUYEN, 2018** | 1- The proposed model (LSSVM-FA based on Least Squared Support Vector Machines (LSSVM) and Firefly algorithm (FA)) performs well with the training data and the validation Since the proposed model is better than benchmarks i.e. Neuron-fuzzy, support vector machines, and random forest,<br>2- The proposed model is a promising tool that should be used for flood modeling use planning and management at flood-prone areas. |
| **9-JOE MARLOU A. OPELLA, 2019** | 1-Fusing ConvNet, a feedforward neural networks that specialize in image processing and prediction with SVM, a supervised machine learning for classification and regression analysis for a better image map results.<br>2-Distinct image prediction output from dilated convolution and deconvolution network will be used as an input to SVM in producing its final output. |
| **10-TALHA AHMED KHAN, 2019** | Comparative analysis has been performed between ML algorithms to determine the highest accuracy algorithm. Parametric comparison and results of training and testing proved that Support Vector Machine (SVM) performed very well. Indeed, Ml classification have been implemented to classify the true positive event of flash floods. |
| **11-MAHYAT SHAFAPOUR TEHRANY, 2019** | 1-SVM produced the highest accuracies of prediction the final susceptibility mapping result, respectively, using DS1 (the LiDAR dataset).<br>2-Additional factors in the modeling, does not necessarily guarantee the achievement of greater accuracy however, the modeling method, can significantly alter outcomes. |
| **12-JOE MARLOU A, 2019** | 1-The SVM also which have the capability in predicting with non-linear data to produce reliable flood susceptibility and probability map.<br>2- The fusion of these two distinct network (CNN-SVM) architecture will produce an effective and robust flood map. |
| **13-GANG ZHAO, 2019** | 1-The proposed weakly labeled support vector machine (WELLSVM), model successfully identified the relationship between the explanatory factors and the urban flooding records, and it outperformed the LR, ANN, and SVM models to Identify flood-prone areas with limited flood inventories.<br>2-WELLSVM showed a high degree-of-fit and prediction ability in the training and testing datasets according to the evaluation indices and ROC curve.<br>3-The high-quality WELLSVM flood susceptibility map is thus applicable to efficient urban flood management. |
| **14-MEIHONG MA,2019** | LSSVM with Radial basis function (RBF) Kernel works the best in terms of accuracy to assess the flash flood risk. |
| **15-ROMULUS COSTACHE, 2020** | 1-Among all of hybrids models the SVM-IoE (SVM- Index of Entropy (IoE)), model has the highest performance for identification of areas susceptible to floods, followed by the FAHP- IoE, IoE, SVM and FAHP.<br>2-Results highlight a very high efficiency of all the applied models.<br>3-Identification of areas are high and very high exposure to these hydrological hazards |
| **16-KE XIONG, 2020** | SVM model capable of producing accurate and robust debris flow susceptibility maps which can provide helpful data for assessing and mitigating debris flow hazards. |

3. The potential of ANNs algorithm for FRA/FRM and producing map to identify flood prone area.

| **REFERENCE** | **ADVANTAGES** |
|---|---|
| **1-JAN ADAMOWSKI, 2012** | 1- The WA-ANN models were found to provide more accurate urban water demand forecasts than the MLR, MNLR, ARIMA, and ANN models<br>2- Coupled wavelet-neural network models are a potentially promising new method of urban water demand forecasting that merit further study. |
| **2-MASOUD BAKHTYARI KIA, 2012** | Integration of artificial neural network (ANN) model and MATLAB software:<br>1-Satisfactory agreement between the predicted and the real hydrological records.<br>2- Help local and national government plan for the future and develop appropriate new infrastructure to protect the lives and property of the people of Johor.<br>3- Model can be used to predict floods in the study area with acceptable accuracy.<br>4- Integration of this model with a real-time warning system will provide a great advantage and flood damages can be reduced significantly. |
| **3-JUNFEI CHEN, 2012** | Combination of Diffusion Theory (DT) and Back propagation (BP) neural network:<br>1- Verification the model's effectiveness for flood risk management.<br>2-This method can be easily applied to effectively resolve problems of insufficient samples in flood risk assessment.<br>3- The information diffusion method is capable of extracting useful information and therefore improves the accuracy of risk assessment. |

| | |
|---|---|
| **4-QIONG LI, 2013** | 1-New methodology (Combination of ANN, Back-Propagation (BP), Information Diffusion Method (IDM) and Fuzzy method) is effective and practical, with the potentiality to be used to forecast flood risk. 2- This method provides an enhanced implementation of the information diffusion process which better corresponds to the actual situation. |
| **5-DIEU TIEN BUI, 2016** | 1-Experimental results showed that the proposed model (Combination of neural fuzzy inference system and Metaheuristic optimization) has high performance on both the and validation dataset, 2- The proposed MONF model outperforms in comparison with the various ML methods such as DT, SVM, RF, MLP, ANF. 3-MONF model is a new alternative tool that should be used in flood susceptibility mapping also it is useful for planners and decision makers for sustainable management of flood-prone areas |
| **6-SEYED VAHID RAZAVI TERMEH, 2018** | Adaptive neuro-fuzzy inference systems (ANFIS) with different metaheuristics algorithms: 1-ant colony optimization (ACO), 2-genetic algorithm (GA), 3- particle swarm optimization (PSO), 4-Learning vector quantization (LVQ), 5-Frequency Ratio (FR) showed: 1-The ANFIS-PSO was found to be the most practical model in term of producing the highly focused flood susceptibility map with lesser spatial distribution related to highly susceptible classes. 2-The results of this study can be used as a platform for better land use planning in order to manage the highly susceptible zones to flooding and reduce the anticipated losses. |
| **7-D. VITO,2018** | 1-An ANN is able to handle non-linearity and automatically adjusts to new information, while generally requiring little computational effort. 2-ANNs are widely accepted as powerful ways of modelling complex non-linear and dynamical systems for which there are large amounts of sometimes noisy data, 3-ANN can thus be trained without having intimate knowledge of the hydrological or other aspects of flood forecasting, 4- SVM can represent a potential alternative which relieves the user from a time consuming trial and error procedure of synthesizing the suitable network architecture as in ANN without compromising on the prediction accuracy. |
| **8-JOE MARLOU A, 2019** | 1-Taking advantage of the capability of convolutional neural networks (CNN) to be effective with high spatial resolution imagery. 2- The SVM also which have the capability in predicting with non-linear data. 3- The fusion of these two distinct network (Conventional Neural Network (CNN) and Support Vector Machine (SVM)) architecture will produce an effective and robust flood map. |
| **9-JOE MARLOU A. OPELLA, 2019** | Exploitation the data available from the GIS and the technology advancement to produce a reliable flood susceptibility and probability map (Conceptual framework) showed: 1-Fusing ConvNet, a feedforward neural networks that specialize in image processing and prediction with SVM, a supervised machine learning for classification and regression analysis for a better image map results. 2-Distinct image prediction output from dilated convolution and deconvolution network will be used as an input to SVM in producing its final output. |
| **10-DIEU TIEN BUI, 2020** | DLNN performs better in comparison with the other methods (MLP, NN, SVM) attains a good prediction accuracy for flash flood susceptibility mapping with Classification Accuracy Rate = 92.05%, Positive Predictive Value = 94.55% and Negative Predictive Value = 89.55%. 2-Proposed hybridization of GIS and deep learning can be a promising tool to assist the government authorities and involving parties in flash flood mitigation and land-use planning. |
| **11-A. NALLAPAREDDY, 2020** | Feed-Forward Neural Network and Cascade-forward back-propagation neural network methods: 1-The Feed-Forward Neural Network gives better accuracy than the Cascade-forward back propagation neural network. 2-Based on the promising results, the proposed method may assist in our understanding of the role of machine learning in disaster detection. |
| **12-A SALEH, 2020** | Weight of Evidence, Random Forest, Analytical Hierarchy Process (AHP), Weight of Evidence (WoE), Frequency Ratio (FR), Artificial Neural Networks (ANN) methods: 1-The GIS coupling method can improve the accuracy of flood susceptibility modelling, the ability of integration between GIS and analytical model shows the good result in developing flash flood susceptibility model. 2- The ensemble and hybrid model of machine learning (combination of models that mentioned above) can increase the accuracy of result. 3- The result of performance ML method has a higher accuracy than qualitative and quantitative method (WoE, FR, and AHP), 4-It is proved that the ability of integration between GIS and analytical model is able to generate the flash flood susceptibility modelling which are very useful tools to plan and manage the flood. |

4. The potential of LR algorithm for FRA/FRM and producing map to identify flood prone area.

| REFERENCE | ADVANTAGES |
|---|---|
| **1-MAHYAT SHAFAPOUR TEHRANY, 2013** | 1-The validation results showed that, area under the curve for the results of DT and integrated method of FR and LR was 87% and 90% for success rate and 82% and 83% for prediction rate respectively.<br>2-DT is based on the rules which are created precisely and strongly by considering all the characteristics of the variables which can enhance the performance of the flood susceptibility mapping.<br>3-FR was used to analyze the impact of classes of each variable on flood occurrence and overcome the weakness of LR |
| **2-MEIHONG MA, 2019** | Application of Artificial Intelligence (Least squares support vector machine: LSSVM) and classical canonical method (Logistic regression: LR) showed:<br>1-LSSVM with Radial basis function (RBF) Kernel works the best in terms of accuracy and the **LR is the worst.**<br>2- The high-risk areas are mainly concentrated in the central and southeastern regions, where with a large curve number |
| **3- MAHFUZUR RAHMAN, 2019** | New approach that have ability to elucidate better identification of flood hazard area. By applying (ANN, LR, FR, AHP) and result showed:<br>1- LR model had the highest success rate (AUROC-81.60%) and prediction rate (86.80%), among other models.<br>2- Integrated LR-FR model had the highest predictive power with an AUROC value of 88.10%.<br>3- A new opportunity for planning and designing flood control measures |
| **4-MASAHIKO HARAGUCHI, 2019** | To illustrate how developing countries with limited datasets and capacity can utilize global hazard data to support risk-informed decision-making at the local level, Visualization, Decision Tree(DT) and Logistic Regression(LR) applied and the results showed:<br>1- Globally available data used with machine learning techniques can be effective for local flood management.<br>2-The best predictive models are found to be:<br>(i) One that uses rainfall type and rainfall amount; (ii) and one that integrates all the information including, rainfall amount, rainfall type and vegetation index. |
| **5-CHINH LUUA,2019** | Combination of Multiple Linear Regression(MLR) and TOPSIS showed**:**<br>1-Identification of regions in Vietnam are at most risk of flood.<br>2- Provides accessible risk information for decision-makers and planners,<br>3- Provides a new method for analyzing Vietnam's national disaster database, can be suggested potential applications for disaster loss databases in other countries and regions. |
| **6-HAMID DARABI, 2020** | Ensemble approach of ML algorithms (1- Boosted Regression Tree (BRT), 2- Multivariate Adaptive Regression Spline (MARS), 3- Generalized Linear Model (GLM), 4-Generalized Additive Model (GAM)) showed:<br>1-Increasing the accuracy of training and testing by ensemble approach ML algorithms, instead of single methods,<br>2-Identification most important factors of flood risk and vulnerability,<br>3-Identification of areas are at most risk of flood leading implement flood risk reduction measures |
| **7-KE XIONG, 2020** | 1-All four models (LR, SVM, RF, BRT) were capable of producing accurate and robust debris flow susceptibility maps.<br>2-the BRT model outperformed other models and was the ideal choice.<br>3-The importance of selecting suitable mapping units and optimal predictors.<br>4- The debris flow susceptibility maps of the Sichuan Province were produced, which can provide helpful data for assessing and mitigating debris flow hazards |

5. The potential of RF algorithm for FRA/FRM and producing map to identify flood prone area.

| REFERENCE | ADVANTAGES |
|---|---|
| **1-ZHAOLI WANG, 2015** | RF algorithm is used to overcome some problems such as complex and systematic process involving high-dimension and non-linear data in flood risk assessment.<br>Both the training and testing error rate of RF can be reduced by increasing the sample size and the number of classification trees.<br>2- The higher and highest-risk zones occupy approximately 19.09% of the total,<br>3-Identification of top five most important causative factors of flood out of the eleven risk indices,<br>4-Provide a novel and highly successful approach (combination of RF and SVM) to flood hazard risk assessment and management for prevention, and reduction of natural disasters in the study basin. |
| **2-CHENGGUANG LAI, 2016** | 1-Ant-Miner exhibits higher accuracy and more simple rules that can be used to generate flood risk zoning map quickly and easily than decision tree method (DT); compared to random forest (RF) and |

| | |
|---|---|
| | fuzzy comprehensive evaluation (FCE), Ant-Miner has significant advantages both in implementation step-reducing and computing time-saving.<br>2-Provide a novel and successful approach to flood risk zoning.<br>3-Evaluation results provide a reference for flood risk management, prevention, and reduction of natural disasters in the study basin. |
| **3-OMID RAHMATI, 2017** | 1-EBF and BRT models performed better than RF model to Identify the Flood prone area (FPA).<br>**CHALLENGS** |
| **4-SUNMIN LEEA, 2017** | The Random forest and Boosted Tree models used to GIS to produce flood susceptibility maps showed:<br>1-The distance from the river, geology and digital elevation model showed a high importance among the other causative flood factors.<br>2-The random forest model showed higher validation accuracies for the regression and classification algorithms, respectively, in comparison with boosted-tree model.<br>3-The flood-susceptibility maps provide meaningful information for decision-makers regarding the identification of priority areas for flood-mitigation management. |
| **5-ALAA M. AL. ABADI, 2018** | The application of ensemble machine Learning classifiers (AdaBoost, random forest and rotation forest model) showed:<br>1-AdaBoost classifier was the best in terms of the statistical measures, followed by the random forest and rotation forest models.<br>2-A flood susceptibility map was prepared based on the result of each classifier and classified in to five zone survey low, low, moderate, high and very high.<br>3-For the model with the best performance, i.e., the AdaBoost model, these zones were distributed over an area (44%) for the very low–low zone, (18%) for the moderate zone and (40%) for the high–very high zones.<br>4-This study proved the high capabilities of ensemble machine learning classifiers to decipher flood susceptibility zones in an arid region. |
| **6-JUNFEI CHEN,2019** | The application of Random Forest (RF) algorithm and Radial Basis Function RBF for accurate assessment of flood risk for regional development illustrated:<br>1-These results were basically in line with the actual situation, which validated the effectiveness of the model.<br>2-Ccounter measures and suggestions for reducing the urban flood risk in the YRD region were proposed, in order to provide decision support for flood control, disaster reduction and emergency management in the YRD region.<br>3- The flood risk has been increasing in the YRD urban agglomeration during the past ten years under the urbanization background, and economic development status showed a significant positive correlation with flood risks. |
| **7-XIANGNAN LI, 2019** | 1-random forest model performed best in comparison of the other methods (LR, NB, and AB) for prediction of global flood susceptibility map.<br>2- The global flood susceptibility map provides a reference for global flood management. |
| **8-BAHRAM CHOUBIN, 2020** | 1-The three ML models (Random Forest (RF), Bagged Classification and Regression Trees (Bagged CART), Mixture Discriminate Analysis (MDA)) achieve an excellent performance (Accuracy > 87% and precision > 86%).<br>2-The Bagged CART and RF models have the same performance and higher than the MDA model.<br>3-Identification of high hazardous areas in Barcelona to predict Spatial hazard maps |
| **9-HAMID REZA POURGHASEMI, 2020** | 1-The results of the multi-hazard map from application of RF that coupled to R software in Shiraz City indicate that 25% of Shiraz city is very susceptible to flooding,<br>2- RF model coupled to R software could increase the accuracy to produce flood susceptibility map |

6. The potential of FR algorithm for FRA/FRM and producing map to identify flood prone area.

| REFERENCE | ADVANTAGES |
|---|---|
| **MAHYAT SHAFAPOUR TEHRANY, 2013** | 1-The validation results showed that, area under the curve for the results of DT and integrated method of FR and LR was 87% and 90% for success rate and 82% and 83% for prediction rate respectively.<br>2-DT is based on the rules which are created precisely and strongly by considering all the characteristics of the variables which can enhance the performance of the flood susceptibility mapping.<br>3-FR was used to analyze the impact of classes of each variable on flood occurrence and overcome the weakness of LR. |
| **MAHYAT SHAFAPOUR TEHRANY, 2015** | 1- Accuracies were increased using the ensemble method (Novel ensemble method: Integrating support vector machine (SVM) and frequency ratio (FR)).<br>2- Efficiency of the proposed ensemble method as rapid, accurate and reasonable in flood susceptibility assessment. |
| **HOSSEIN MOJADDADI,2017** | 1-The ensemble method of FR and SVM can be efficiently used in flood hazard studies because of its simple structure and robust performance.<br>2-The FR model is an excellent approach for ranking different classes of conditioning parameters. |

| | |
|---|---|
| | 3- Derived map can be helpful to planners and decision makers for flood management |
| **A SALEH, 2020** | 1-The GIS coupling method can improve the accuracy of flood susceptibility modelling, the ability of integration between GIS and analytical model shows the good result in developing flash flood susceptibility model.<br>2- The ensemble and hybrid model of machine learning can increase the accuracy of result.<br>3- The result of performance ML method (Weight of Evidence, Random Forest, Analytical Hierarchy Process (AHP), Weight of Evidence (WoE), Frequency Ratio (FR), Artificial Neural Networks (ANN)) has a higher accuracy than qualitative and quantitative method (WoE, FR, and AHP).<br>4-It is proved that the ability of integration between GIS and analytical model is able to generate the flash flood susceptibility modelling which are very useful tools to plan and manage the flood. |

7. The number of research papers in last decades that used single ML method and conventional data for FRA

| Modeling Technique | Reference | Flood Resource Variables | Case study |
|---|---|---|---|
| **1-Random Forest (RF)**<br>**2-Rotation Forest**<br>**3-AdaBoost** | Alaa M. Al-Abadi, 2018 | 1- Flood inventory map<br>2-10 flood conditions factors:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics | Iraq |
| **1- Logistic Regression (LR),**<br>**2-Naive Bayes (NB),**<br>**3-Ada Boost (AB),**<br>**4-Random Forest (RF)** | Xiangnan Li, 2019 | 1-Flood inventory map<br>2-13 condition factor of flood:<br>2.1-Rainfall<br>2.2-Topographic data | Global/ Fourth level watersheds |
| **1-Random Forest (RF)**<br>**2-Radial Basis Function (RBF)** | Junfei Chen,2019 | 1-Season rainfall,<br>2-urban impervious area ratio,<br>3-gross domestic product (GDP) per square kilometer of land,<br>4-water area ratio,<br>5-population density | Yangtze River Delta (YRD), China |

8. The number of research papers in last decades that used single ML method, combined conventional and less conventional data for FRA

| Modeling Technique | Reference | Flood Resource Variables | Case study |
|---|---|---|---|
| **Machine Learning classification algorithm** | Leonardo B. L. Santos**,** 2017 | 1- Historical flood inundated areas,<br>2-Remote Sensing,<br>3- Hydrological Data,<br>4- Health Data,<br>5- Demography Data,<br>6-Transportation Data | Madeira River, Western Brazilian Amazon |

9. The number of research papers in last decades that used hybrid ML methods and conventional data for FRA

| Modeling Technique | Reference | Flood Resource Variables | Case study |
|---|---|---|---|
| **Diffusion Theory (DT) combined with Back propagation (BP) neural network** | Junfei CHEN, 2012 | Five risk evaluation indexes:<br>1-density of population (M),<br>2-density of industrial and agricultural products (N),<br>3-density of one-line embankment (W),<br>4-average difference between control points embankment top elevation and every year's highest water level (S),<br>5-proportion of danger embankment length to the total embankment length (T). | Dongting Lake, China |
| **1-Artificial Neural Network (ANN, Back-Propagation (BP))**<br>**2-Information Diffusion Method (IDM),**<br>**3-Fuzzy method(F)**<br>**4-Combind; BP-IDM-F** | Qiong Li,2013 | 1-Disaster area,<br>2-Dead Population,<br>3-Collapsed houses<br>4-Disaster degree | China |

| | | | |
|---|---|---|---|
| **1- Novel ensemble method: Integrating support vector machine (SVM) and frequency ratio (FR).<br>2- Decision tree(DT)** | Shafapour Tehrany, M. 2015 | 1-Flood inventory map,<br>2-10 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Kelantan basin, Malaysia |
| **1-Random Forest (RF) combined with<br>2- Support Vector Machine(SVM)** | Zhaoli Wang,2015 | 1-Flood inventory map,<br>2-11 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Dongjiang River Basin, China |
| **1- Artificial intelligence approach based on neural fuzzy inference system,<br>2-Metaheuristic optimization<br>2.1-Evolutionary Genetic,<br>2.2- Particle Swarm Optimization.<br>3-New method from combined 2 previous method MONF** | Dieu Tien Bui, 2016 | 1- Historical flood inundated areas<br>2-10 flood conditions factors:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Tuong Duong, Vietnam |
| **1- A hybrid model based on<br>1.1-ASD (Automated regression-based Statistical Downscaling tool)<br>1.2-KNN (K-nearest neighbor)<br>2-Bayesian neural network (BNN),<br>3- KNN<br>4-Generalized Extreme Value (GEV)** | Y. Lu,2016 | 1-Historical Rainfall<br>2- Meteorological data<br>3-Hydrometric data | Duhe River, China |
| **1-Ant Colony Algorithm, combine with Fuzzy Comprehensive Evaluation (FCE),<br>2-Decision Tree (DT)<br>3-Random Forest (RF)** | Chengguang Lai, 2016 | -Flood inventory map,<br>2-14 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Dongjiang River, China |
| **LSSVM-FA based on Least Squared Support Vector Machines (LSSVM) and Firefly algorithm (FA)** | Viet-Nghia Nguyen, 2018 | 1- Historical flood inundated areas<br>2-10 flood conditions factors:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Central Vietnam |
| **1-Artificial Intelligence (Least squares support vector machine: LSSVM with Radial basis function (RBF) Kernel).<br>2-classical canonical method (Logistic regression: LR)** | Meihong Ma, 2019 | 1- Historical flash flood records<br>2-13 flood conditions factor:<br>2.1-Topographic data<br>2.2-Anthropological<br>2.3-Hydrological<br>2.4-Rainfall data | Yunnan, China |
| **1-Vise kriterijumska optimizacijaik ompromisno Resenje (VIKOR) integration with (TOPSIS),<br>3-Simple Additive Weighting (SAW)<br>4-Naïve Bayes Tree (NBT)<br>5-Naïve Bayes(NB)** | Khabat Khosravi, 2019 | 1-Flood inventory map,<br>2-12 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Ningdu Catchment. China |
| **Multiple Linear Regression(MLR) combined with TOPSIS** | Chinh Luua,2019 | 11 Causative factors for disaster lossless such as: human loss, housing, education, health Care and so on. | Vietnam |
| **Combined MCDM & ML methods:<br>1-Multi-Criteria Decision-Making (MCDM) analysis techniques<br>1.1-VIKOR,** | Khabat Khosravi, 2019 | 1-Flood inventory map,<br>2-12 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Ningdu Catchment/china |

| | | | |
|---|---|---|---|
| **1.2-TOPSIS**<br>**1.3-SAW**<br>**2- machine learning methods**<br>**2.1-Bayes Tree (BT),**<br>**2.2-Naïve Bayes(NBT))** | | | |
| **1-A hybrid physics-machine learning approach: Long Short-Term Memory Units (LSTM)**<br>**2-Five global hydrological models (GHMs) under the Inter-Sectoral Impact Model Inter-comparison Project 2b (ISMIP2b)** | Yang, T, 2019 | 1032 gauging stations data in the Global Streamflow Indices and Metadata Archive (GSIM) for the period of 1971-2010. | Global |
| **Combined ML & statistical models**<br>**1-Logistic Regression, Classification (LRC- CART)**<br>**2-Regression Trees (RT),**<br>**3-Multi Layer Perceptron (MLP),**<br>**4-Random Forest (RF)**<br>**5-Support Vector Machine (SVM)**<br>**6-Decision Tree (DT)** | Romulus Costache, 2020 | 1-8 flood conditions factor:<br>1.1-Topographic data<br>1.2-geological characteristics<br>1.3-soil characteristics | Bâsca Chiojdului River, Romania |
| **Combined ML & F;**<br>**1-Multi-Layer Perceptron(MLP),**<br>**2- weighted k-nearest neighbors (wkNN),**<br>**3-Fuzzy-weighted k-nearest neighbors (FwkNN)** | Mohd Hafizul Afifi Abdullah, 2020 | Rainfall data | Kuala Krai, Malaysia |

10. The number of research papers in last decades that used hybrid ML methods, combined conventional and less conventional data for FRA

| **Modeling Technique** | **Reference** | **Flood Resource Variables** | **Case study** |
|---|---|---|---|
| **Hybridization of:**<br>**1-The artificial neural network (ANN) model**<br>**2-MATLAB software** | Masoud Bakhtyari Kia, 2012 | 1-Flood inventory map,<br>2-7 flood conditions factor:<br>2.1-Topographic data<br>2.2-Rainfall data<br>2.3-Remote Sensing data<br>2-4 Survey from public | Johor River Basin, Malaysia |
| **Hybridization of:**<br>**1-Visualization,**<br>**2-Decision Tree(DT)**<br>**3-Logistic Regression(LR)** | Masahiko Haraguchi, 2019 | 1--Rainfall data:<br>1-1- Rain gauge data<br>1.2-Remote sending data | Metro Manila |
| **Hybridization of ML & DST:**<br>**1-Boosted generalized linear model (GLMBoost) 2-Random Forest (RF),**<br>**3-Bayesian generalized linear model (BayesGLM))** | Farzaneh Sajedi Hosseini, 2020 | 1--Flood inventory map,<br>2-13 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data<br>2.5-Satillite data | Gorganroud River, Iran |
| **Ensemble approach: ML algorithms:**<br>**1- Boosted Regression Tree (BRT),**<br>**2- Multivariate Adaptive Regression Spline (MARS),**<br>**3- Generalized Linear Model (GLM),**<br>**4-Generalized Additive Model (GAM)** | Hamid Darabi, 2020 | 1.8 flood conditions factor:<br>1.1-Topographic data<br>1.2-geological characteristics<br>1.3-soil characteristics<br>1.4-Rainfall data<br>1.5-Public Survey | Amol, Iran |

| | | | |
|---|---|---|---|
| **Hybridization of ML & AHP:**<br>**1-Weight of Evidence, Random Forest,**<br>**2-Analytical Hierarchy Process (AHP),**<br>**3-Weight of Evidence (WoE),**<br>**4-Frequency Ratio (FR),**<br>**5-Artificial Neural Networks (ANN).** | A Saleh, 2020 | Geographic Information System (GIS) coupling with empirical observation-model | Global |

11. The number of research papers in last decades that used single ML method and conventional data to produce reliable flood map

| **Modeling Technique** | **Reference** | **Flood Resource Variables** | **Case study** |
|---|---|---|---|
| **1. Evidential belief function (EBF)**<br>**2. Random forest (RF)**<br>**3. Boosted regression trees (BRT)** | Omid Rahmati, 2017 | 1-Flood inventory map<br>2-10 condition factors of flood:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics | Galikesh, Iran |
| **1-Random Forest (RF),**<br>**2-Boosted Tree models (BT)** | Sunmin Leea,2017 | 1-- Flood inventory map<br>2-12 flood conditions factors:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics | Seoul, Korea |
| **1- Alternating Decision Tree (ADT),**<br>**2-Functional Tree (FT),**<br>**3-Kernel Logistic Regression (KLR),**<br>**4-Multi Layer Perceptron (MLP),**<br>**5-Quadratic Discriminant Analysis(QDA)** | Saeid Janizadeh, 2019 | 1-Historical flood events<br>2- 8 flood conditions factor:<br>2.1-Topographic data<br>2.2-Rainfall data | Tafresh, Iran |
| **1-weakly labeled support vector machine (WELLSVM),**<br>**2- logistic regression (LR),**<br>**3-artificial neural networks (ANNs),**<br>**4- support vector machine(SVM)** | Gang Zhao, 2019 | 1-Flood inventory map,<br>2-9 flood conditions factor:<br>2.1-Topographic data<br>2.2-Rainfall data<br>2.3-geological characteristics<br>2.4- Anthropogenic | Beijing, China |
| **1-Kernel Logistic Regression(KLR),**<br>**2-Radial Basis Function Classifier (RBFC),**<br>**3-Multinomial Naïve Bayes (NBM),**<br>**4-Logistic Model Tree (LMT)** | Binh Thai Pham,2020 | 1-Flood inventory map,<br>2-10 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics | Nghe An Vietnam |
| **1-Deep Learning Neural Network (DLNN)) algorithm,**<br>**2-Multilayer Perceptron (MLP),**<br>**3-Neural Network (NN)**<br>**4-Support Vector Machine (SVM)** | Dieu Tien Bui, 2020 | 1--Flood inventory map,<br>2-9 flood conditions factor:<br>2.1-Topographic data<br>2.2-geological characteristics<br>2.3-soil characteristics<br>2.4-Rainfall data | Vietnam |

12. The number of research papers in last decades that used single ML method and less conventional data to produce reliable flood map

| **Modeling Technique** | **Reference** | **Flood Resource Variables** | **Case study** |
|---|---|---|---|
| **1-- Supervised method with all data (a supervised neural network approach),**<br>**2-Supervised method with optical data only (a support vector machine (SVM) classifier)** | Nathan ongbotham,2012 | Images from two separates satellite:<br>1-(SPOT)<br>2-(ERS-1) | Gloucester, UK |

| | | | |
|---|---|---|---|
| **3- Unsupervised method with optical data only (an unsupervised clustering algorithm),**<br>**4- A predictive model (digital elevation model (DEM))** | | | |
| **1-Feed-Forward Neural Network,**<br>**2-Cascade-forward back-propagation neural network** | A. Nallapareddy,2020 | Image data from satellite | Uttar Pradesh, India |

13. The number of research papers in last decades that used single ML method, conventional and less conventional data to produce reliable flood map

| **Modeling Technique** | **Reference** | **Flood Resource Variables** | **Case study** |
|---|---|---|---|
| **1-Decision Tree (DT)**<br>**2-Support Vector Machine (SVM)** | Mahyat Shafapour Tehrany, 2019 | 1-Flood inventory map,<br>2-13 flood conditions factor:<br>2.1-Topographic data<br>2.3-Remote Sensing data from 2 center<br>2.4-geological characteristics<br>2.5-soil characteristics | Brisbane, Australia |
| **1-Logistic Regression(LR),**<br>**2-Support Vector Machines(SVM),**<br>**3-Random Forest(RF),**<br>**4-Boosted Regression Trees (BRT),** | Ke Xiong, 2020 | 1-Flood inventory map<br>2.4 flood conditions factor:<br>altitude, altitude difference, aridity index, and groove gradient<br>2.4-Remote sensing data | Sichuan, China |

## 8. Validation Methods

Validation methods such as Area Under the ROC Curve (AUC) and various statistical measures were used to validate and compare the models in this study. ROC curve is a popular measure to evaluate the accuracy of the model and can be used to determine the accuracy of natural hazard susceptibility mapping [37, 38, 40, 110, 111, 112]. Two values are used to build the ROC curve: sensitivity and 100- specificity [113, 114, 96, 79, 80, 81]. Performance of the models is analyzed quantitatively using the area under the curve (AUC) [82-90],[93,94]. An AUC value of 1 indicates the best classification, while 0.5 corresponds to non-accurate models [93,94]. AUC values are calculated according to the equation:

$$\mathrm{AUC} = \sum \mathrm{TP} + \sum \frac{\mathrm{TN}}{\mathrm{P}} + \mathrm{N} \qquad (6)$$

Where TP and TN are considered the rate of pixels classified correctly as flood and non-flood, P and N are the total number of flash floods and non-flash floods, respectively. Various statistical measures such as accuracy (ACC), sensitivity (SST), specificity (SPF), root mean squared errors (RMSE), kappa (K) positive predictive value (PPV), and negative predictive value (NPV) were also selected to validate flood flash modelling [94]. PPV and NPV are the values of pixel probabilities classified correctly as "flood" occurrence and "non-flood" occurrence [95]. The proportion of flash flood pixels is represented by SST value and proportion of non-flash flood pixels is represented by SPF. K is used to analyze the accuracy of modelling [88]. K value varies between -1 and 1. Values of K close to 1 represent better reliability [7]. ACC is the ratio of the rate number of correct predictions and the total number of predictions [98]. RMSE represents the difference between data observations and data estimates [116-123]. Equations for the different measures are given below:

$$\mathrm{SST} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (7)$$

$$\mathrm{SPF} = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \qquad (8)$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{FP+TP}} \qquad (9)$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{FN+TN}} \qquad (10)$$

$$\mathrm{K} = \frac{\mathrm{P_p - P_{exp}}}{1 - \mathrm{P_{exp}}} \qquad (11)$$

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(X_{predicted} - X_{actual})^2} \quad (13)$$

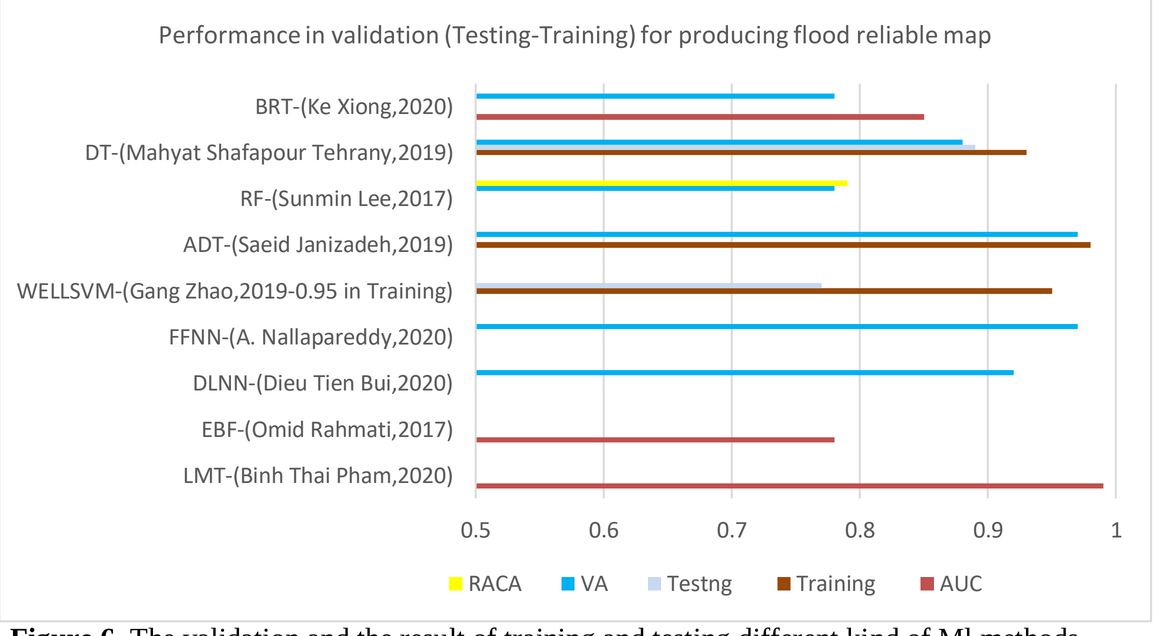


**Figure 6.** The validation and the result of training and testing different kind of Ml methods in recent decades (Single and combined ML method to produce reliable flood maps)

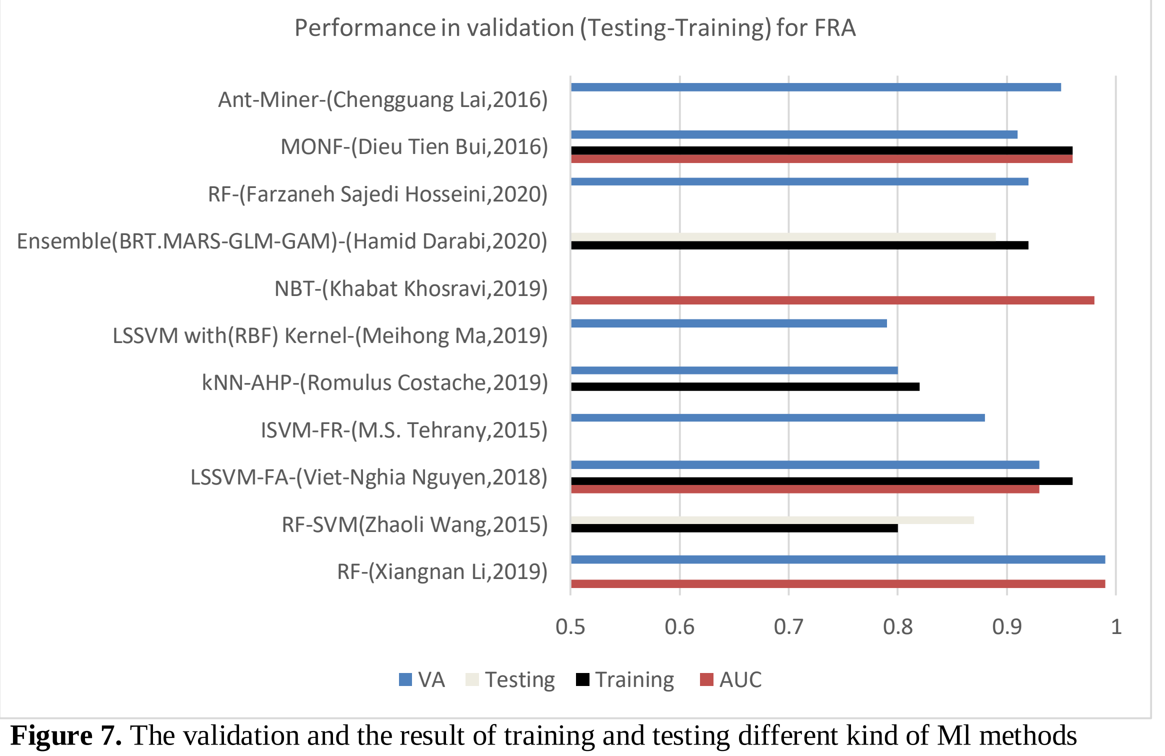


**Figure 7.** The validation and the result of training and testing different kind of Ml methods in recent decades (Single and combined ML methods for flood risk assessment)

Where FP and FN are the rate of pixels classified incorrectly as the flood and non-flood. $\boldsymbol{P_p}$ is the rate of pixels classified correctly for flood or non-flood. Expected agreements is defined by $\boldsymbol{P_{exp}}$. $X_{predicted}, X_{actual}$ are the predicted and real values in the training samples or the testing samples of the models, and n is the total number of samples in the training samples or testing samples. [27].
In figure 6 and 7 the performance of different kinds of ML methods are shown in recent decades.

## 9. Open issues and future research directions

Based on the detailed survey reported in previous sections, we further indicate a number of open issues and suggest some future research directions.
First, the machine learning methods used for data fusion are simplex. As we discussed in previous sections, most of machine learning models mentioned for data fusion are based on SVM, clustering and neural networks, which are classical methods and simple neural networks. SVM and clustering methods often aim at classifying with high accuracy. NN is suitable for describing uncertain complex systems. Nevertheless, the power of machine learning methods should be far more than this. Taking one example, deep learning is considered as a significant research field in artificial intelligence in next 10 years. Deep learning describes the techniques that simulate complex neural systems of humans. Compared with simple neural networks, more hidden layers inserted into the network would give the system better accuracy and learning quality. The lack of deep learning methods for data fusion motivate us to explore new thoughts.
Second, researchers pay little attention to fusion efficiency of Ml method for FRA and producing reliable flood map. Refer to section 7, past work focuses more on fusion quality than fusion efficiency. Some works even did not discuss or evaluate this important property at all. The most obvious disadvantage of machine learning methods is its computational complexity and huge consumption of computing and system resources. Machine learning often needs large sets of data for training, which also brings difficulty into actual applications. Since there will be a good deal for specific needs of miniature devices in the future, which are not affordable for complicated computation due to limited resources, the study for optimizing the efficiency of data fusion models becomes necessary.
Third, comprehensive concern of data fusion by Ml method which is aimed at FR and producing flood reliable map is missed. Based on Table 1, few literatures discussed Robustness and Extensibility. Some literatures did not testify if their models are stable in an unsteady environment with experimental results. These requirements should be fundamental for a fusion model. Some works consider little about the models' effectiveness in practical use. Taking Robustness as an example, data with serious imprecision, inconsistency and noises often occurred, a model that cannot handle this circumstance well will be practically limited. A similar argument is put on Extensibility. Simply improving data fusion accuracy and quality, but ignoring other properties will lead to an imperfect model, while a comprehensive model that satisfies all expected criteria should be urgently studied.
Finally, few existing literatures take account of data scares environment combined conventional data and less conventional data can be effective on accuracy? how will be the computational efficiency? Machine learning methods have a great need to deal with a large scale

of data sets to ensure learning quality and fusion accuracy. However, using original data in machine learning could cause sensitive information leakage. This problem can be particularly acute in the Internet related applications such as intrusion detection, attack analysis, and location tracking. Private information about identities and positions of data providers could be disclosed if the proposed model cannot manage it well.
Another future research direction is the use of more complex and large-scale learning techniques into data fusion via ML methods. As talked above, we place expectations on deep learning, which combines supervised learning and unsupervised learning to construct learning hierarchy, namely the network. Especially in some scenarios that relate to a large amount of data, Deep learning can gain much more improved performance and prediction precision than past learning algorithms [37]. According to [4], there have been some efficient models appeared to deal with fusion problems with deep learning. In [125], a deep belief network based data fusion scheme was proposed for ball screw fault detection. Nevertheless, there might be some following challenges introduced at the same time. The effectiveness of deep learning can only be ensured with mass data and high resource consumption. How to ensure the applicability of deep learning based fusion models in small devices and how to make trade-off between fusion efficiency and quality are additional issues that should be solved. Except for the issues mentioned above, we are also looking forward to researches on deep composite intelligent applications.
Advances in hydrology, meteorology, engineering, using of GIS and remote sensing still not able to increase real time forecast. Researchers from developed countries have stressed to more focus to improve very short time an effective early warning system with collaboration of local communities for flood risk supervision.
The other issue need to improve researches is, in meteorology, flash flood forecasting and lead time early warning still one of the challenging task [124].

## 11. Conclusion:

The current state of ML modeling for flood risk assessment is quite young and in the early stage of advancement. This literature review presents an overview of ML models used in FRA, and develops a classification scheme to analyze the existing literature. The survey represents the performance analysis and investigation of more than 1000 articles. Among them, we identified 110 original and influential articles where the performance and accuracy of at least two machine learning models were compared. To do so, the assessment models were classified into three categories according to type of ML methods, hybrid and single methods, and further divided into type of data (single data or fusion conventional and less conventional data). The state of the art of these classes was discussed and analyzed in detail, considering the performance comparison of the methods available in the literature. The performance of the methods was evaluated in terms of AUC and RMSE, in addition to the generalization ability, robustness, computation cost, and speed. Despite the promising results already reported in implementing the most popular machine learning methods, e.g. SVM, DTs, ANNs, LR and RF there was significant research and experimentation for further improvement and advancement. In this context, there were four major trends reported in the literature for improving the quality of FRA. The first was novel hybridization, either through the integration of two or more machine learning methods or the integration of a machine learning method(s) with more conventional means, and/or soft computing. The second was the use of data fusion techniques for the purpose of improving the quality of the dataset, which highly contributed in improving the accuracy of FRA. The third was the use of an ensemble of methods, which dramatically increased the generalization ability of the models and decreased the uncertainty of FRA. The fourth was the use of add-on optimizer algorithms to improve the quality of machine learning algorithms, e.g., for better tuning the DT or SVM to reach optimal classification or architectures. It is expected that, through these four key technologies, FRA will witness significant improvements, also producing map for identification flood prone areas. Surely, the advancement of these novel ML methods depends highly on the proper usage of soft computing techniques in designing novel learning algorithms. This fact was discussed in the literature review, and the soft computing techniques were introduced as the main contributors in developing hybrid ML methods of the future. Here, it is also worth mentioning that the multidisciplinary nature of this work was the most challenging difficulty to overcome in this research. Having contributions from the coauthors of both realms of ML and hydrology was the key to success. Furthermore, the novel search methodology and the creative taxonomy and classification of the ML methods led to the original achievement of the research. For future work, conducting a survey on spatial FRA using machine learning models is highly encouraged. This important aspect of FRA was excluded from our paper due to the nature of modeling methodologies and the datasets used in assessment the location of floods. Nevertheless, the recent advancements in machine learning models for spatial flood analysis revolutionized this particular realm of flood forecasting, which requires separate investigation. Indeed, current research provides a concise and comprehensive reference for researchers and practitioners in the field of FRA via ML methods. Nevertheless, the literature still lacks a thorough review of the recent advances of machine learning and data fusion technique for FRA/FRM. Therefore, it is beneficial to review and summarize the state of the art in order to gain a deep insight on how machine learning can benefit and optimize data fusion in order to flood risk assessment and producing flood reliable map.

## Appendix

Logistic Model Tree (LMT)
Evidential belief function (EBF)
Deep Learning Neural Network (DLNN)
Weakly labeled support vector machine (WELLSVM)
Alternating decision tree (ADT)
Random Forest (RF)
Decision Tree (DT)
Boosted Regression Trees (BRT)
Validation Accuracy (VA)
Regression and classification algorithms (RACA)
Neural fuzzy inference system and metaheuristic optimization for flood susceptibility modeling, namely MONF
LSSVM-FA) based on Least Squared Support Vector Machines (LSSVM) and Firefly algorithm (FA)
integrating support vector machine (SVM) and frequency ratio (FR)=(ISVM-FR)
Least squares support vector machine: LSSVM with Radial basis function (RBF) Kernel
Boosted regression tree, multivariate adaptive regression spline, generalized linear model, and generalized additive model(BRT.MARS-GLM-GAM)
Dempster–Shafer theory (DST)
CD (Conventional Data)
LCD (Less Conventional Data)
Single(S)
Combined(C)
Flood Risk Assessment (FRA)
Optical System Probatoire d'Observation de la Terre (SPOT) and
SAR instrument aboard the European Remote Sensing (ERS-1)
Fuzzy (F)
Combined Machine Learning and Mathematical Modeling(C-ML-M)
Combined machine learning, and multi-criteria decision analysis(C-ML&MCDA)
Area under the receiver operating curve (AUROC)
Regression Technique (RT)
Conventional neural networks (CNN)
Logistic Model Tree (LMT).
Logistic Regression (LR),
Bayesian Logistic Regression (BLR),
Random Forest (RF)
Decision Tree (DT)
Evidential belief function (EBF)
Boosted regression trees (BRT)
Naive Bayes (NB)
Ada Boost (AB)
Multiple Linear Regression (MLR),
Multiple Non-Linear Regression (MNLR),
Auto Regressive Integrated Moving Average (ARIMA),
Aartificial Neural Networks (ANNs)
Wavelet transforms (WA)
Support Vector Machine (SVM)
Digital Elevation Model (DEM)
Frequency Ratio (FR)
Least Squares Support Vector Machine: LSSVM
Genetic Algorithm Rule-Set Production (GARP)
Quick Unbiased Efficient Statistical Tree (QUEST)
Fuzzy Analytical Network Process (FANP)
Vise kriterijumska Optimizacijaik ompromisno Resenje (VIKOR),
Simple Additive Weighting (SAW)
Naïve Bayes Tree (NBT)
Logistic Model Trees (LMT),
Reduced Error Pruning Trees (REPT),
Alternating Decision Trees (ADT)
Weakly Labeled Support Vector Machine (WELLSVM)
Alternating Decision Tree (ADT)
Functional Tree (FT)
Kernel Logistic Regression (KLR),
Multi-Layer Perceptron (MLP),
Quadratic Discriminant Analysis (QDA)
Analytical Hierarchy Process (AHP)
Back Propagation neural network (BP)
Information Diffusion Method (IDM)

### Nomenclatures

| | |
|---|---|
| WMO | World meteorological organization |
| GCM | Global circulation models |
| SPOTA | Seasonal Pacific Ocean temperature Analysis |
| ANN | artificial neural network |
| POTA | Pacific Ocean temperature analysis |
| QPE | Quantitative precipitation estimation |
| CLIM | Climatology average method |
| EOF | Empirical orthogonal function |
| MLR | Multiple linear regressions |
| QPF | Quantitative precipitation forecasting |
| MNLR | Multiple nonlinear regressions |
| ML | Machine learning |
| MLR | Multiple linear regression |
| ANN | Neural networks |
| WNN | Wavelet-based neural network |
| ARIMA | Autoregressive integrated moving average |
| USGS | United States Geological Survey |
| FFA | Flood frequency analyses |

| | |
|---|---|
| QRT | Quantile regression techniques |
| SPOTA | Seasonal Pacific Ocean temperature analysis |
| SVM | Support vector machines |
| LS-SVM | Least-square support vector machines |
| AI | Artificial intelligence |
| VRM | Vector Regression Machine |
| FFNN | Feed-forward neural network |
| FBNN | Feed-backward networks |
| MLP | Multilayer perceptron |
| ANFIS | Adaptive neuro-fuzzy inference system |
| BPNN | Backpropagation neural network |
| SVR | Support vector regression |
| LR | Linear regression |
| FIS | Fuzzy inference system |
| CART | Classification and regression tree |
| LMT | Logistic model trees |
| NWP | Numerical weather prediction |
| NBT | Naive Bayes trees |
| ARMA | Autoregressive moving averaging |
| REPT | Reduced-error pruning trees |
| DT | Decision tree |
| ELM | Extreme learning machine |
| EPS | Ensemble prediction systems |
| SNIP | Source normalized impact per paper |
| SRM | Structural risk minimization |
| SJR | SCImago journal rank |
| ARMAX | Linear autoregressive moving average with exogenous input |
| LMT | logistic model trees |
| ARMA | Autoregressive moving averaging |
| ADT | Alternating decision trees |
| NARX network | Nonlinear autoregressive network with exogenous inputs |
| RMSE | Root-mean-square error |
| RFFA | Regional flood frequency analysis |
| NLR | Nonlinear regression |
| AR | Autoregressive |

| | |
|---|---|
| WARM | Wavelet autoregressive model |
| NLR-R | Nonlinear regression with Regionalization approach |
| E | Nash Sutcliffe index |
| FR | Frequency ratio |
| SOM | Self-organizing map |
| CHIM | Cluster-based hybrid inundation model |
| FFRM | Flash flood routing model |
| KGE | Kling-Gupta efficiency |
| AME | ANN-based monsoon rainfall enhancement |
| SSNN | State-space neural network |
| SSL | Suspended sediment load |
| NSE | Nash–Sutcliffe efficiency |
| E-CHAID | Exhaustive CHAID |
| CHAID | Chi-squared automatic interaction Detector |
| CLIM | climatology average Model |
| HECX-HMS | Hydrologic engineering left-hydrologic modeling system |
| SOM | Self-organizing map |
| PBIAS | Percent bias |
| NLPM | Nonlinear perturbation model |
| RF | Rotation forest |
| KSOFM-NNM | Kohnen self-organizing festure maps neural networks model |
| DBPANN | divsion-based backpropagation DBP neural network |
| NLPM-ANN | Nonlinear perturbation model basedon neural network |
| GRNNM | Generalized regression neural networks model |
| IIS | Iterative input selection |
| EEMD | Ensemble empirical mode decomposition |
| ANNE | Artificial neural network ensembles |
| DWT | Discrete wavelet transform |
| SFF | Seasonal flood forecasting |
| MP | Water monitoring points |
| WBANN | Wavelet–bootstrap–ANN |
| HBI | Hilsenhoff’s biotic index |
| RT | Regression trees |

| | |
|---|---|
| EMD | Empirical mode |
| decomposition | |
| LLR | Local linear regression |
| BFGS | Broyden Fletcher Goldfarb |
| Shanno | |
| M-EMD | Modified empirical mode |
| | decomposition IIS |
| SAR | Seasonal first-order |
| autoregressive | |
| BFGSNN | Broyden Fletcher Goldfarb |
| Shanno | |
| | neural network |
| GRNN | Artificial neural networks |
| including | |
| | Generalized regression |
| network | |
| T-S | Takagi-Sugeno |
| WLGP | Wavelet linear genetic |
| programming | |
| E | Nash coefficients |
| TSC-T–S | Clustering based Takagi– |
| Sugeno | |
| TCs | Tropical cyclones |
| PCA | Principal component |
| analysis | |

**References**


1. 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.

2. Xie, K.; Ozbay, K.; Zhu, Y.; Yang, H. Evacuation zone modeling under climate change: A data-driven method. *J. Infrastruct. Syst.* 2017, *23*, 4017013.

3. Lohani, A.K.; Goel, N.; Bhatia, K. Improving real time flood forecasting using fuzzy inference system. *J. Hydrol.* 2014, *509*, 25–41.
4. Mosavi, A.; Bathla, Y.; Varkonyi-Koczy, A. Predicting the Future Using Web Knowledge: State of the Art Survey. In *Recent Advances in Technology Research and Education*; Springer: Cham, Switzerland, 2017; pp. 341–349.

5. Zhao, M.; Hendon, H.H. Representation and prediction of the indian ocean dipole in the poama seasonal forecast model. *Q. J. R. Meteorol. Soc.* 2009, *135*, 337–352.

6. Borah, D.K. Hydrologic procedures of storm event watershed models: A comprehensive review and comparison. *Hydrol. Process.* 2011, *25*, 3472–3489.

7. Costabile, P.; Costanzo, C.; Macchione, F. A storm event watershed model for surface runoff based on 2D fully dynamic wave equations. *Hydrol. Process.* 2013, *27*, 554–569.
8. Cea, L.; Garrido, M.; Puertas, J. Experimental validation of two-dimensional depth-averaged models for forecasting rainfall–runoff from precipitation data in urban areas. *J. Hydrol.* 2010, *382*, 88–102.

9. Fernández-Pato, J.; Caviedes-Voullième, D.; García-Navarro, P. Rainfall/runoff simulation with 2D full shallow water equations: Sensitivity analysis and calibration of infiltration parameters. *J. Hydrol.* 2016, *536*, 496–513.

10. Caviedes-Voullième, D.; García-Navarro, P.; Murillo, J. Influence of mesh structure on 2D full shallow water equations and SCS curve number simulation of rainfall/runoff events. *J. Hydrol.* 2012, *448*, 39–59.

11. Costabile, P.; Costanzo, C.; Macchiono, F. Comparative analysis of overland flow models using finite volume schemes. *J. Hydroinform.* 2012, 14, 122.

12. Xia, X.; Liang, Q.; Ming, X.; Hou, J. An efficient and stable hydrodynamic model with novel source term discretization schemes for overland flow and flood simulations. *Water Resour. Res.* 2017, *53*, 3730–3759.
13. Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land

surface water and energy fluxes for general circulation models. *J. Geophys. Res. Atmos.* 1994, *99*, 14415–14428.

14. Costabile, P.; Macchione, F. Enhancing river model set-up for 2-D dynamic flood modeling. *Environ. Model. Softw.* 2015, 67, 89–107.

15. Nayak, P.; Sudheer, K.; Rangan, D.; Ramasastri, K. Short-term flood forecasting with a neurofuzzy model. *Water Resour. Res.* 2005, *41*.

16. Kim, B.; Sanders, B.F.; Famiglietti, J.S.; Guinot, V. Urban flood modeling with porous shallow-water equations: A case study of model errors in the presence of anisotropic porosity. *J. Hydrol.* 2015, *523*, 680–692.

17. Bellos, V.; Tsakiris, G. A hybrid method for flood simulation in small catchments combining hydrodynamic and hydrological techniques. *J. Hydrol.* 2016, *540*, 331–339.

18. Bout, B.; Jetten, V. The validity of flow approximations when simulating catchment-integrated flash floods. *J. Hydrol.* 2018, *556*, 674–688.

19. Costabile, P.; Macchione, F.; Natale, L.; Petaccia, G. Flood mapping using lidar dem. Limitations of the 1-D modeling highlighted by the 2-D approach. *Nat. Hazards* 2015, *77*, 181 –204.

20. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. *J. Math. Stat.* 2012, *8*, 330–338.

21. Adamowski, J.; Fung Chan, H.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour.Res.* 2012, *48*.

22. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* 2013, *476*, 433–441.

23. Chow, V.T.; Maidment, D.R.; Larry, W. *Mays. Applied hydrology*; International Edition; MacGraw Hill, Inc.: New York, NY, USA, 1988; p. 149.

24. Aziz, K.; Rahman, A.; Fang, G.; Shrestha, S. Application of artificial neural networks in regional flood frequency analysis: A case study for australia. *Stoch. Environ. Res. Risk Assess.* 2014, *28*, 541 554.

25. *Kroll, C.N.; Vogel, R.M. Probability distribution of low streamflow series in the united states. J. Hydrol. Eng. 2002, 7, 137–146.*

26. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Spatial prediction of flood susceptible areas using rule based

decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J. Hydrol.* 2013, *504*, 69–79.

27. Romulus Costache, Mihnea Cristian Popa,Dieu Tien Bui, Daniel Constantin Diaconu, Nicu Ciubotaru, Gabriel Minea, Quoc Bao Pham. Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *Journal of Hydrology*.2020, Volume 585, 124808.

28. Ke Xiong, Basanta Raj Adhikari, Constantine A. Stamatopoulos, Yu Zhan, Shaolin Wu, Zhongtao Dong, Baofeng Di. Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China. *Remote Sens*. 2020, 12, 295; doi:10.3390/rs12020295

29. Huang, S.; Chang, J.; Huang, Q.; Chen, Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J. Hydrol.* 2014, *511*, 764–775.

30. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.* 2014, *512*, 332–343.

31. Liong, S.Y.; Sivapragasam, C. Flood stage forecasting with support vector machines. *JAWRA J. Am. Water Resour. Assoc.* 2002, *38*, 173– 186.

32. Lafdani, E.K.; Nia, A.M.; Pahlavanravi, A.; Ahmadi, A.; Jajarmizadeh, M. Research article daily rainfall-runoff prediction and simulation using ANN, ANFIS and conceptual hydrological MIKE11/NAM models. *Int. J. Eng. Technol.* 2013, *1*, 32–50.

33. Mekanik, F.; Imteaz, M.; Gato-Trinidad, S.; Elmahdi, A. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *J. Hydrol.* 2013, *503*, 11–21.

34. Mosavi, A.; Rabczuk, T.; Varkonyi-Koczy, A.R. Reviewing the novel machine learning tools for materials design. In *Recent Advances in Technology Research and Education*; Springer: Cham, Switzerland, 2017; pp. 50–58.

35. Xu, Z.; Li, J. Short-term inflow forecasting using an artificial neural network model. *Hydrol. Process.* 2002, *16*, 2423–2439.

36. Kim, S.; Matsumi, Y.; Pan, S.; Mase, H. A real-time forecast model using artificial neural network for after-runner storm surges on the Tottori Coast, Japan. *Ocean Eng.* 2016, *122*, 44–53.

37. Grecu, M.; Krajewski, W. A large-sample investigation of statistical procedures for radar-based short-term quantitative precipitation forecasting. *J. Hydrol.* 2000, *239*, 69–84.

38. Maddox, R.A.; Zhang, J.; Gourley, J.J.; Howard, K.W. Weather radar coverage over the contiguous united states. *Weather Forecast.* 2002, *17*, 927–934.

39. Gizaw, M.S.; Gan, T.Y. Regional flood frequency analysis using support

vector regression under historical and future climate. *J. Hydrol.* 2016, *538*, 387–398.

40. Campolo, M.; Andreussi, P.; Soldati, A. River flood forecasting with a neural network model. *Water Resour. Res.* 1999, *35*, 1191–1197.

41. Saeid Janizadeh, Mohammadtaghi Avand, Abolfazl Jaafari, Tran Van Phong, Mahmoud Bayat , Ebrahim Ahmadisharaf , Indra Prakash, Binh Thai Pham, Saro Lee. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. *Sustainability* 2019, 11, 5426;

42. Binh Thai Pham, Tran Van Phong, Huu Duy Nguyen, Chongchong Qi, Nadhir Al-Ansari, Ata Amini, Lanh Si Ho, Tran Thi Tuyen, Hoang Phan Hai Yen, Hai-Bang Ly, Indra Prakash, Dieu Tien Bui. *Water* 2020, 12, 239;

43. Joe Marlou A. Opella. Alexander A. Hernandez Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework. *IEEE 15th International Colloquium on Signal Processing & its Applications (CSPA 2019),* 2019.

44. Jan Adamowski, Hiu Fung Chan, Shiv O. Prasher, Bogdan Ozga-Zielinski,and Anna Sliusarieva. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resource Research*. 2012. 48, W01528.

45. Seyed Vahid Razavi Termeh, Aiding Kornejady, Hamid Reza Pourghasemi, Saskia Keesstra. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of The Total Environment.* 2018, 615, 438- 451.

46. A.Saleh, A. Yuzir, I. Abustan. Flash Flood Susceptibility Modelling: A Review. *OP Conf. Series: Materials Science and Engineering* 712 (2020).

47. Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. *J. Hydrol.* 2015, *527*, 1130–1141.

48. Chengguang Lai, Quanxi Sha, Xiaohong Chen, Zhaoli Wang, Xiaowen Zhou, Bing Yang, Lilan Zhang, Flood risk zoning using a rule mining based on ant colony algorithm. *Journal of Hydrology.* 2016, Volume 542, 268-280.

49. Hamid Reza pourghasemi, narges Kariminejad, Mahdis Amiri, Mohsen edalat, Mehrdad Zarafshar, thomas Blaschke, Artemio Cerda. Assessing and mapping multi hazard risk susceptibility using a machine learning technique. *Scientific Reports*. (2020), 10:3203.

50. Mahyat Shafapour Tehrany, Biswajeet Pradhan, Mustafa Neamah Jebur. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*. 2013, Volume 504, 69- 79.

51. Chinh Luu, Jason von Meding, Mohammad Mojtahedi. Analyzing Vietnam's national disaster loss

database for flood risk assessment using multiple linear regression-TOPSIS. *International Journal of Disaster Risk Reduction.*2019

52. Hossein Mojaddadi, Biswajeet Pradhan, Haleh Nampak, Noordin Ahmad, Abdul Halim bin Ghazali. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Natural Hazards and Risk.*2017 8(2)1080-1102.

53. De'ath, G.; Fabricius, K.E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 2000, *81*, 3178 3192.

54. Dehghani, M.; Saghafian, B.; Rivaz, F.; Khodadadi, A. Evaluation of dynamic regression and artificial neural networks models for real-time hydrological drought forecasting. *Arabian J. Geosci.* 2017, *10*, 266.

55. Choubin, B.; Khalighi-Sigaroodi, S.; Malekian, A.; Ahmad, S.; Attarod, P. Drought forecasting in a semi-arid watershed using climate signals: A neuro-fuzzy modeling approach. *J. Mt. Sci.* 2014, *11*, 1593 1605.

56. Choubin, B.; Khalighi-Sigaroodi, S.; Malekian, A.; Ki ̧si, Ö. Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzyinferencesystemforforecastingprecipitationbasedonlarge-scaleclimate signals. *Hydrol. Sci. J.* 2016, *61*, 1001–1009.

57. Farzaneh Sajedi Hosseini, Bahram Choubin, Amir Mosavi, Narjes Nabipoure, Shahaboddin Shamshirband, Hamid Darabi, Ali Torabi Haghighi. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of The Total Environment*. 2020, 711, 135161.

58. Tsai, L.T.; Yang, C.-C. Improving measurement invariance assessments in survey research with missing data by novel artificial neural networks. *Expert Syst. Appl.* 2012, *39*, 10456–10464.

59. Mahyat Shafapour Tehrany, Biswajeet Pradhan, Mustafa Neamah Jebur. Flood susceptibility mapping using a novel ensemble weights- of evidence and support vector machine models in GIS. *Journal of Hydrology*, 2014. (5)12, 332-343

60. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 2016, *13*, 361–378.

61. Etemad-Shahidi, A.; Mahjoobi, J. Comparison between m5$^J$ model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Eng.* 2009, *36*, 1175–1181.

62. Tan, Q.-F.; Lei, X.-H.; Wang, X.; Wang, H.; Wen, X.; Ji, Y.; Kang, A.-Q. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. *J. Hydrol.* 2018.

63. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* 1998, *13*, 18–28.

64. Vapnik, V.; Mukherjee, S. Support vector method for multivariate density estimation. *Adv. Neural Inf. Process. Syst.* 2000, *4*, 659–665.

65. Li, S.; Ma, K.; Jin, Z.; Zhu, Y. A new flood forecasting model based on SVM and boosting learning algorithms. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 1343 1348.

66. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model induction with support vector machines: Introduction and applications. *J. Comput. Civ. Eng.* 2001, *15*, 208–216. [CrossRef]

67. Nayak, M.A.; Ghosh, S. Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. *Theor. Appl. Climatol.* 2013, *114*, 583–603.

68. Ortiz-García, E.; Salcedo-Sanz, S.; Casanova-Mateo, C. Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. *Atmos. Res.* 2014, *139*, 128–136.

69. Granata, F.; Gargano, R.; de Marinis, G. Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA's storm water management model. *Water* 2016, *8*, 69. [CrossRef]

70. Gong, Y.; Zhang, Y.; Lan, S.; Wang, H. A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near lake okeechobee, Florida. *Water Resour. Manag.* 2016, *30*, 375–391.

71. Jajarmizadeh, M.; Lafdani, E.K.; Harun, S.; Ahmadi, A. Application of SVM and swat models for monthly streamflow prediction, a case study in South of Iran. *KSCE J. Civ. Eng.* 2015, *19*, 345–357.

72. Bao, Y.; Xiong, T.; Hu, Z. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing* 2014, *129*, 482–493.

73. Bray, M.; Han, D. Identification of support vector machines for runoff modelling. *J. Hydroinform.* 2004, *6*, 265–280.

74. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* 2015, *125*, 91–101.

75. Kisi, O.; Parmar, K.S. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J. Hydrol.* 2016, *534*, 104–112.

76. Sachindra, D.; Huang, F.; Barton, A.; Perera, B. Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *Int. J. Climatol.*2013, *33*, 1087–1106.

77. Choubin, B.; Zehtabian, G.; Azareh, A.; Rafiei-Sardooi, E.; Sajedi-

Hosseini, F.; Ki ̦si, Ö. Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. *Environ. Earth Sci.* 2018, *77*, 314.

78. Abbot, J.; Marohasy, J. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmos. Res.* 2014, *138*, 166–178.

79. Li, L.; Xu, H.; Chen, X.; Simonovic, S. Streamflow forecast and reservoir operation performance assessment under climate change. *Water Resour. Manag.* 2010, *24*, 83.

80. *Wu, C.; Chau, K.-W. Data-driven models for monthly streamflow time series prediction. Eng. Appl. Artif. Intell. 2010, 23, 1350–1367.*

81. Sulaiman, J.; Wahab, S.H. Heavy rainfall forecasting model using artificial neural network for flood prone area. In *It Convergence and Security 2017*; Springer: Singapore, 2018; pp. 68–76.

82. Kar, A.K.; Lohani, A.K.; Goel, N.K.; Roy, G.P. Development of flood forecasting system using statistical and ANN techniques in the downstream catchment of mahanadi basin, india. *J. Water Resour. Prot.* 2010, *2*, 880.

83. Jain, A.; Prasad Indurthy, S. Closure to "comparative analysis of event-based rainfall-runoff modeling techniques—Deterministic, statistical, and artificial neural networks" by ASHU JAIN and SKV prasad indurthy. *J. Hydrol. Eng.* 2004, *9*, 551–553.

84. Lohani, A.; Kumar, R.; Singh, R. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *J. Hydrol.* 2012, *442*, 23–35.

85. Tanty, R.; Desmukh, T.S. Application of artificial neural network in hydrology—A review. *Int. J. Eng. Technol. Res.* 2015, *4*, 184–188.

86. Ki ̦si, O. Streamflow forecasting using different artificial neural network algorithms. *J. Hydrol. Eng.* 2007, *12*, 532–539.

87. Shamseldin, A.Y.Artificial neural network model for river flow forecasting in a developing country. *J. Hydroinform.* 2010, *12*, 22–35.

88. Badrzadeh, H.; Sarukkalige, R.; Jayawardena, A. Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting. *J. Hydrol.* 2013, *507*, 75–85.

89. Smith, J.; Eli, R.N. Neural-network models of rainfall-runoff process. *J. Water Resour. Plan. Manag.* 1995, *121*, 499–508.

90. Taormina, R.; Chau, K.-W.; Sethi, R. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice Lagoon. *Eng. Appl. Artif. Intell.* 2012, *25*, 1670 1676.

91. Fotovatikhah, F.; Herrera, M.; Shamshirband, S.; Chau, K.-W.; Faizollahzadeh Ardabili, S.; Piran, M.J. Survey of computational intelligence as basis to big flood management: Challenges, research

directions, and future work. *Eng. Appl. Comput. Fluid Mech.* 2018, *12*, 411–437.

92. Faizollahzadeh Ardabili, S.; Najafi, B.; Alizamir, M.; Mosavi, A.; Shamshirband, S.; Rabczuk, T. Using SVM-RSM, and ELM-RSM Approaches for Optimizing the Production Process of Methyl and Ethyl Esters. *Energies* 2018, *11*, 2889.

93. Thirumalaiah, K.; Deo, M. River stage forecasting using artificial neural networks. *J. Hydrol. Eng.* 1998, *3*, 26–32.

94. *Panagoulia, D. Artificial neural networks and high and low flows in various climate regimes. Hydrol. Sci. J. 2006, 51, 563–587.*

95. Panagoulia, D.; Tsekouras, G.; Kousiouris, G. A multi-stage methodology for selecting input variables in ann forecasting of river flows. *Glob. Nest J.* 2017, *19*, 49–57.
96. Zadeh, M.R.; Amin, S.; Khalili, D.; Singh, V.P. Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions. *Water Resour. Manag.* 2010, *24*, 2673– 2688.

97. Deo, R.C.; S ̧ahin, M. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in Eastern Australia. *Atmos. Res.* 2015, *161*, 65–81.
98. Coulibaly, P.; Dibike, Y.B.; Anctil, F. Downscaling precipitation and temperature with temporal neural networks. *J. Hydrometeorol.* 2005, *6*, 483–496.

99. Schoof, J.T.; Pryor, S. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *Int. J. Climatol.* 2001, *21*, 773–790.

100. Zhang, J.-S.; Xiao, X.-C. Predicting chaotic time series using recurrent neural network. *Chin. Phys. Lett.* 2000, *17*, 88.

101. *Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. Neurocomputing 2006, 70, 489–501.*

102. Lima, A.R.; Cannon,A.J; Hsieh, W. W. Forecasting daily stream flow using online sequential extreme learning machines. *J. Hydrol.* 2016, *537*, 431–443.

103. Yaseen, Z.M.; Jaafar, O.; Deo, R.C.; Kisi, O.; Adamowski, J.; Quilty, J.; El-Shafie, A. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq. *J. Hydrol.* 2016, *542*, 603–614.

104. Mahyat Shafapour Tehrany, Biswajeet Pradhan, Shattri Mansor, Noordin Ahmad. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* . 2019, Volume 125, 91-101.

105. Mahfuzur Rahman, Chen Ningsheng, Md Monirul Islam, Ashraf Dewan, Javed Iqbal, Rana Muhammad Ali Washakh, Tian Shufen. Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-Criteria Decision Analysis. *Earth Systems and Environment*. 2019, 3:585–601.

106. Bahram Choubina, Ehsan Moradi, Mohammad Golshan, Jan Adamowski, Farzaneh Sajedi-Hosseini, Amir Mosavi. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment*. 2019, Volume 651, Part 2, 2087-2096.

107. Xiangnan, Li.; Denghua, Yan. ; Kun, Wang. ; Baisha, Weng. ; Tianling Qin. ; Siyu Liu. Flood Risk Assessment of Global Watersheds Based on Multiple Machine Learning Models, *Water Journal*.2019, 11, 1654.

108. Mosavi, A.; Edalatifar, M. A.; Edalatifar, M. A Hybrid Neuro-Fuzzy Algorithm for Prediction of Reference Evapotranspiration. In *Recent Advances in Technology Research and Education*; Springer: Cham, Switzerland, 2018; pp. 235–243.

109. Dineva, A.; Várkonyi-Kóczy, A.R.; Tar, J.K. Fuzzy expert system for automatic wavelet shrinkage procedure selection for noise suppression. In Proceedings of the 2014 IEEE 18th International Conference on Intelligent Engineering Systems (INES), Tihany, Hungary, 3–5 July 2014; pp. 163 -168.

110. Prakash, O.; Sudheer, K.; Srinivasan, K. Improved higher lead time river flow forecasts using sequential neural network with error updating. *J. Hydrol. Hydromech.* 2014, *62*, 60–74.

111. Shu, C.; Ouarda, T. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* 2008, *349*, 31–43.

112. Ashrafi, M.; Chua, L.H.C.; Quek, C.; Qin, X. A fully-online neuro-fuzzy model for flow forecasting in basins with limited data. *J. Hydrol.* 2017, *545*, 424–435.

113. Yu, P.-S.; Yang, T.-C.; Chen, S.-Y.; Kuo, C.-M.; Tseng, H.-W. Comparison of random forests and support vector machine for real-time radar- derived rainfall forecasting. *J. Hydrol.* 2017, *552*, 92–104.

114. Nourani, V.; Baghanam, A.H.; Adamowski, J.; Kisi, O. Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. *J. Hydrol.* 2014, *514*, 358–377. [CrossRef]

115. Hassan, Z.; Shamsudin, S.; Harun, S.; Malek, M.A.; Hamidon, N. Suitability of ANN applied as a hydrological model coupled with statistical downscaling model: A case study in the northern area of peninsular Malaysia. *Environ. Earth Sci.* 2015, *74*, 463–477.

*116.* Sahoo, G.; Ray, C. Flow forecasting for a Hawaii stream using rating curves and neural networks. J. Hydrol. 2006, 317, 63–80.

117. Kim, S.; Singh, V.P.Flood forecasting using neural computing techniques and conceptual class segregation. *JAWRA J. Am. Water Resour. Assoc.* 2013, *49*, 1421–1435.

118. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. Nature 1986, 323, 533.

119. Riad, S.; Mania, J.; Bouchaou, L.; Najjar, Y. Rainfall-runoff model using an artificial neural network approach. *Math. Comput. Model.* 2004, *40*, 839–846.

120. Senthil Kumar, A.; Sudheer, K.; Jain, S.; Agarwal, P.Rainfall-runoff modeling using artificial neural networks: Comparison of network types. *Hydrol. Process. Int. J.* 2005, *19*, 1277–1291.

121. Zadeh, L.A. Soft computing and fuzzy logic. In *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi a Zadeh*; World Scientific: Singapore, 1996; pp. 796–804.

122. Choubin, B.; Darabi, H.; Rahmati, O.; Sajedi-Hosseini, F.; Kløve, B. River suspended sediment modelling using the cart model: A comparative study of machine learning techniques. *Sci. Total Environ.* 2018, *615*, 272–281.

123. Bogardi, I.; Duckstein, L. The fuzzy logic paradigm of risk analysis. In *Risk-Based Decisionmaking in Water Resources X*; American Society of Civil Engineers: Reston, VA, USA, 2003; 12–22.

124. Sivapalan, M.; Blöschl, G.; Merz, R.; Gutknecht, D. Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resour. Res.* 2005, *41*.

125. Dandagala, S.; Reddy, M.S.; Murthy, D.S.; Nagaraj, G. Artificial neural networks applications in groundwater hydrology—A review. *Artif. Intell. Syst. Mach. Learn.* 2017, *9*, 182–187.