

Review

Open Access



# Intelligent flood forecasting and warning: a survey

Yue Zhang<sup>1</sup>, Daiwei Pan<sup>1</sup>, Jesse Van Griensven<sup>1,2</sup>, Simon X. Yang<sup>1</sup>, Bahram Gharabaghi<sup>1</sup>

<sup>1</sup>School of Engineering, University of Guelph, Guelph ON N1G 2W1, Canada.

<sup>2</sup>Mech and Robotics Engineering, University of Waterloo, Waterloo ON N2L 3G1, Canada.

**Correspondence to:** Prof. Bahram Gharabaghi and Simon X. Yang, School of Engineering, University of Guelph, Guelph ON N1G 2W1, Canada. E-mails: bgharaba@uoguelph.ca; syang@uoguelph.ca

**How to cite this article:** Zhang Y, Pan D, Van Griensven J, Yang SX, Gharabaghi B. Intelligent flood forecasting and warning: a survey. *Intell Robot* 2023;3(2):190-212. <http://dx.doi.org/10.20517/ir.2023.12>

**Received:** 15 Mar 2023 **First Decision:** 21 Apr 2023 **Revised:** 17 May 2023 **Accepted:** 30 May 2023 **Published:** 28 Jun 2023

**Academic Editor:** Chen Peng **Copy Editor:** Yanbing Bai **Production Editor:** Yanbing Bai

## Abstract

Accurately predicting the magnitude and timing of floods is an extremely challenging problem for watershed management, as it aims to provide early warning and save lives. Artificial intelligence for forecasting has become an emerging research field over the past two decades, as computer technology and related areas have been developed in depth. In this paper, three typical machine learning algorithms for flood forecasting are reviewed: supervised learning, unsupervised learning, and semi-supervised learning. Special attention is given to deep learning approaches due to their better performance in various prediction tasks. Deep learning networks can represent flood behavior as powerful and beneficial tools. In addition, a detailed comparison and analysis of the multidimensional performance of different prediction models for flood prediction are presented. Deep learning has extensively promoted the development of real-time accurate flood forecasting techniques for early warning systems. Furthermore, the paper discusses the current challenges and future prospects for intelligent flood forecasting.

**Keywords:** flood forecasting, intelligent prediction, supervised learning, unsupervised learning, semi-supervised learning, deep learning

## 1. INTRODUCTION

Flooding, especially in developing countries, is one major cause of fatalities among both humans and animals. In addition to the loss of life, flooding damages property and destroys crops. Floods can be dangerous to hu-



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



mans and animals, and timely flood forecasting can help prevent or mitigate disasters. Therefore, it is essential to develop prediction models for accurate flood forecasting.

Floods can be defined as the result of water overflowing and submerging land that is normally dry. Flooding is the phenomenon of a rapid increase in the flow of water due to a sudden surge in the volume of water in rivers, lakes, and seas. The main natural factors leading to changes in water volume and levels include heavy rainfall, rapid melt of ice and snow, and storm surges, among others<sup>[1]</sup>. Generally speaking, the direct economic damage caused by floods to human beings is comparable to that of other natural disasters such as earthquakes and hurricanes. Floods are characterized by a wide range of impacts, extreme destructiveness, and rapidity of disaster<sup>[2]</sup>. Floods can be categorized into different types based on their causes and severity, including river floods, flash floods, coastal floods, and urban floods. In addition to these categories, floods can be classified based on their severity or frequency. Depending on meteorological conditions and geographical factors, floods are also characterized by elevated seasonal frequency and regional vulnerability.

The global assessment report of the Intergovernmental Panel on Climate Change (IPCC) on natural disasters shows that 313 natural disasters occurred worldwide in 2020, affecting 123 countries and regions<sup>[3]</sup>. Among them, flood disasters had the highest frequency, occurring 193 times, accounting for 61.66% of the total, and directly affecting a population of 33,215,600<sup>[4]</sup>. The worldwide direct economic losses caused by natural disasters amounted to US\$173 billion, with storms, floods, wildfires, and earthquakes accounting for 99% of the losses. For example, the floods that occurred in Henan, China, in 2020 caused a direct economic loss of US\$17 billion, the costliest natural disaster worldwide that year, causing losses to the urban economy, agricultural development, etc.<sup>[5]</sup>. In some cases, floods can act as triggers for numerous indirect disasters, including geological disasters such as mudslides and landslides, health disasters such as plagues and viruses, and negative impacts on the environment and climate<sup>[6]</sup>. Such secondary disasters generally have long recovery cycles and high recovery costs, and the hazard of some disasters is irreversible.

Real-time prediction of floods is essential to lay the foundation for mitigating damage to human property and planning defenses<sup>[7]</sup>. Datasets and machine learning algorithms are two of the most important factors that influence the accuracy of the forecasting result. Firstly, selecting the raw data for flood forecasting involves a careful and systematic process of identifying relevant variables, gathering historical data, preprocessing the data, selecting relevant features, integrating the data, and splitting it into training and testing sets. Then high-quality datasets are critical for building accurate and effective machine learning models. These datasets should be large, diverse, well-labeled, balanced, clean, representative of the problem, and ethically collected. Real-time flood forecasting uses historical data and information from past events to predict flooding. This is because historical data are easier to collect than real-time data. Once processed, historical data (high-quality datasets) are then used to create models for predicting where and when flooding may occur in the future. Therefore, the warnings can be issued to the public so that they can prepare for floods<sup>[8,9]</sup>. Artificial intelligence (AI) technology has been employed increasingly by governments to create automated flood forecasting systems<sup>[10]</sup>. At first, supervised learning technology is used for forecasting because the labeled inputs are easier to process and achieve good results. As the need for forecasting increases, there is a growing demand to use unlabeled data, leading to the study of unsupervised learning problems. In some cases, forecasting inputs contain both labeled and unlabeled data, for which semi-supervised learning is of particular interest in order to reach better forecasting. It has been observed that in most situations, conventional supervised learning, unsupervised learning, and semi-supervised learning approaches cannot maintain good performance as the forecasting factors involved become complicated, and the size of forecasting features increases. To overcome this challenge, deep learning approaches have received enough attention recently in the area of forecasting. While machine learning methods have made a great contribution to forecasting, it remains to be quite challenging when applying such methods to natural disaster prediction, such as floods, which could be much more complicated. In the past few years, some primary results have been reported in flood prediction by means of a variety of common

machine learning methods. For example, Spatio-Temporal Attention Long Short Term Memory (STA-LSTM) model has great performance in basic flood forecasting, but its robustness and generalization have been improved.<sup>[11]</sup> Nonetheless, with social development and climate change, there has been increasing natural and human factors affecting flood prediction significantly. Such impacts mainly result from rising temperatures, snow melt, ice melt, rainstorms, and soil conditions, and the main human factors include high urbanization, population explosion, and overexploitation of trees. In addition, simple machine learning cannot be suitable for the complex spatio-temporal datasets because natural and human factors are evolving. Furthermore, each flood prediction algorithm has its own characteristics, such as robustness, generalization, calculating speed, gradient problems, weight problems, and fitting problems. Therefore, the purpose of this paper is to provide a comprehensive understanding of the current state of machine learning in flood prediction and to encourage further research in this area.

Based on the above observations, we believe that it is timely and necessary to conduct a review of the current application of machine learning in flood prediction, particularly an abundant literature survey in real-time flood forecasting. While machine learning techniques have been widely applied in various prediction tasks, it remains challenging to obtain stable prediction performance in the area of flood forecasting, especially due to the diversity in the spatio-temporal datasets<sup>[12]</sup>. Given the large amount of literature available, we specifically analyze and compare the performance of mainstream algorithms currently used in real-time flood prediction. Of particular interest is how these algorithms perform when faced with huge spatio-temporal datasets; it is of great importance to understand whether or not they can still maintain high levels of accuracy and robustness. This aspect has not been shown in previous studies. The contributions of this paper are as follows: (1) Several practical problems and popular machine learning methods used for flood prediction are presented, and their advantages and disadvantages with respect to the computational cost, gradient problem, and robustness and accurateness are discussed in depth; (2) The difficulties in the development of a real-time flood prediction are exhaustively analyzed in terms of the characteristic of flood prediction, spatio-temporal data, and noise. Based on this, a complete literature assessment of recent achievements in dealing with these difficulties is done; (3) The results of this study are summarized, and a number of interesting research trajectories are identified that may help to further advancement of this field.

The rest of the paper is organized as follows. Section 2 introduces the machine learning techniques related to flood prediction. Section 3 addresses several common technical issues that arise in machine learning algorithms designed for flood forecasting. Section 4 includes the comparisons of flood prediction models and some valuable challenges and future work. Finally, Section 5 is the conclusion of this paper.

## 2. MACHINE LEARNING FOR FLOOD FORECASTING

The performance of a flood forecasting model is heavily influenced by the prediction algorithm used in the model. This section highlights three common challenges in machine learning and deep learning models and presents a model that is well-suited for handling large amounts of data.

### 2.1. Supervised learning

Supervised learning is a form of AI that can support flood prediction<sup>[13,14]</sup>. Supervised learning algorithms are used to identify flood levels and give early warning. In the forecasting process, these algorithms are a boon in determining how to use historical data most quickly and accurately to successfully predict future disasters<sup>[15-17]</sup>. Supervised learning requires input from the user - meaning that they need to distinctly state what they are looking for before they can effectively identify patterns or make predictions about user input data<sup>[18]</sup>. There are several supervised learning models that can be used to predict flood events: decision trees, k-nearest neighbors, support vector machines (SVM), and neural networks<sup>[19]</sup>. Each model has its advantages and disadvantages, while SVM is the most commonly used method for flood forecasting<sup>[20]</sup>. They are easy

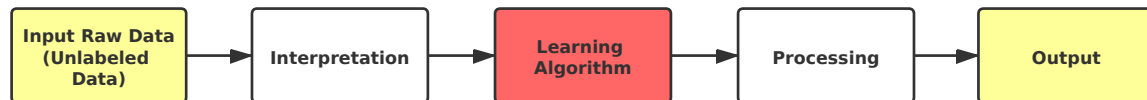


Figure 1. Unsupervised learning block diagram.

to implement and have a high accuracy rate compared to other methods. In addition, they have optimal performance in high-dimensional spaces.

The training dataset for supervised learning has two attributes  $(x_i, y_i)$ , where “ $x$ ” stands for the system supervision input and “ $y$ ” for the system output. Meanwhile, “ $i$ ” is the training sample indicator for “ $x$ ” and “ $y$ ”, respectively. In supervised learning, a training input “ $x_i$ ” is provided to the learning system, which then produces an output “ $\tilde{y}_i$ ”. An arbitrator then calculates the difference between the two outputs for comparison with the ground truth marker “ $y_i$ ”. The discrepancy is referred to as the error signal, which is then transmitted to the learning system to modify the settings of learners. The learning process aims to generate ideal learning system parameters with high accuracy by reducing the difference between “ $\tilde{y}_i$ ” and “ $y_i$ ” for all “ $i$ ” [21,22]. A set of discrete values or a vector space is represented by the input or output. The arbitrator is not subject to any unique limitation under the learning paradigm. “ $y_i$ ” - “ $\tilde{y}_i$ ” is typically used to calculate the error signal when “ $y_i$ ” is taken from a continuous space. The arbitrator typically generates the error signal based on the equality between  $y_i$  and “ $\tilde{y}_i$ ” if “ $y_i$ ” is one of a set of discrete values. The arbitrator assumes output to be 0 for same “ $y_i$ ” and “ $\tilde{y}_i$ ” and 1 for the different “ $y_i$ ” and “ $\tilde{y}_i$ ” [22,23].

Flood prediction models rely on historical data gathered from diverse locations worldwide, including Canada, Pakistan, China, India, and Bangladesh [24]. These models are trained with the collected data to forecast future flood events, enabling preparedness measures in the face of such events. This allows for preparedness when a flood event occurs. The supervised learning algorithm is an effective way of accurately predicting future events based on historical datasets [25]. It is especially useful for protecting lives during dangerous floods and other natural disasters.

## 2.2. Unsupervised learning

Unlike supervised learning, there is no supervision to be given in unsupervised learning; for example, it is usually used to address the datasets that are not labeled. The aim of unsupervised learning is to infer the underlying structure of the datasets provided, in which some evident groups can be identified [26,27].

In unsupervised learning, a computer program learns without any labeled input, as shown in Figure 1. Models for real-time flood prediction rarely incorporate unsupervised learning. This is due to the subjectivity and absence of specific analytical goals, such as response prediction, that characterize unsupervised learning. Evaluation of the results of unsupervised learning techniques is also hard because there is no commonly accepted method for cross-validation or testing the results across several datasets.

## 2.3. Semi-supervised learning

Semi-supervised learning is a type of machine learning based on supervised learning with semi-supervised data. It is an extension of supervised learning and unsupervised learning. In semi-supervised learning, the algorithm learns the model using labeled data and unlabeled data, as shown in Figure 2. Therefore, semi-supervised learning is a useful tool in situations with a shortage of labeled data [28,29]. As such, unlabeled data can help create better classifiers if there are enough data available and certain distributional assumptions about the data are accurate [30,31]. Moreover, since many machine learning applications exist, developers always try to improve the algorithms by developing new ideas. One of these ideas is semi-supervised learning, which makes it easy to train models for different tasks without having access to all the training data.

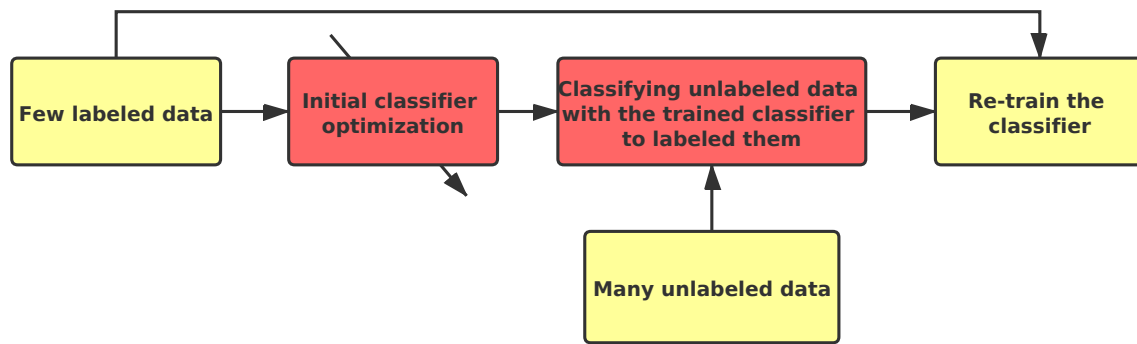


Figure 2. Semi-supervised learning block diagram.

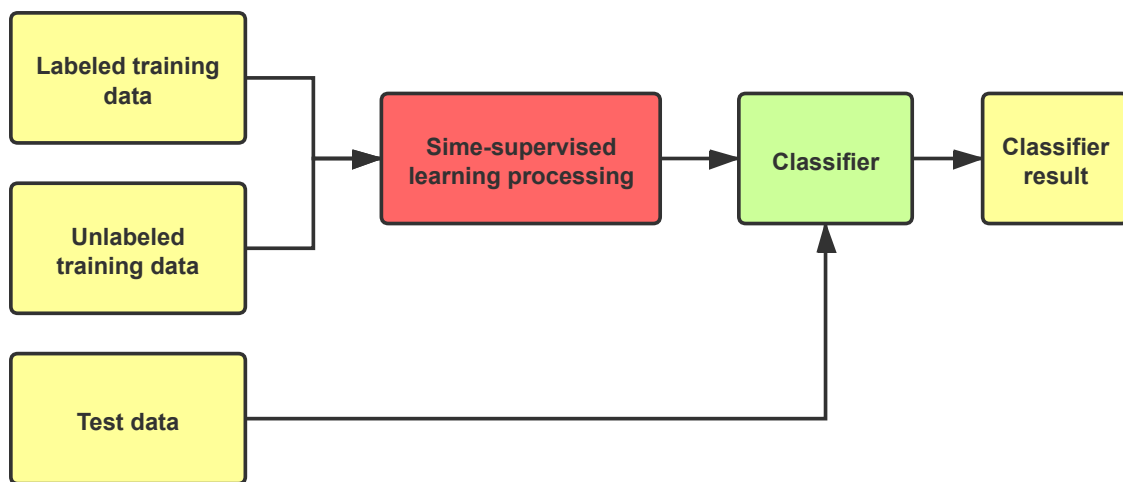
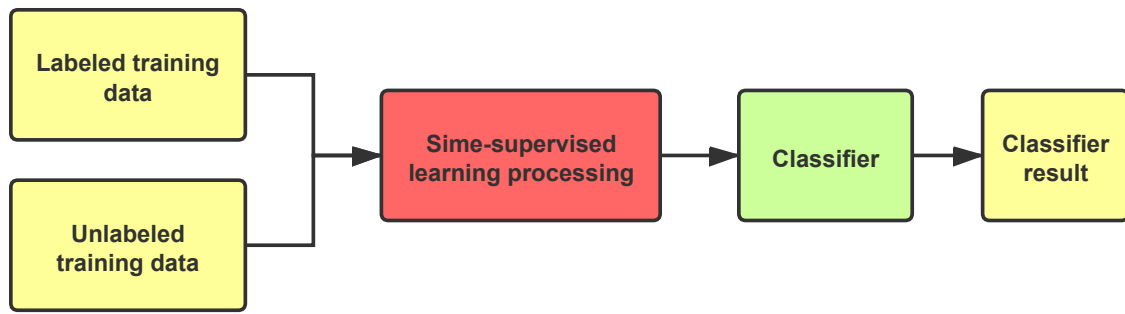


Figure 3. Inductive semi-supervised learning diagram.

According to the statistical learning theory, it can be further divided into inductive semi-supervised learning, as shown in Figure 3 and transductive semi-supervised learning, as shown in Figure 4<sup>[32]</sup>. The full data collection contains two different sorts of sample sets. Let  $D_{Labeled} = \{X_{train}, Y_{train}\}$  denote the labeled sample set and  $D_{Unlabeled} = \{X_{unknown}, X_{test}\}$  the unlabeled sample set, and suppose  $C_{D_{Unlabeled}} \gg C_{D_{Labeled}}$ . For inductive semi-supervised learning, denote  $D_{train} = \{X_{train}, Y_{train}, X_{unknown}\}$  the training set, and  $X_{unknown}$  and  $X_{test}$  are both unlabeled sets and  $X_{unknown} \neq X_{test}$ . For transductive semi-supervised learning,  $D_{train} = \{X_{train}, Y_{train}, X_{unknown}\}$  is the training set, and  $X_{unknown}$  is unlabeled and  $X_{unknown} = X_{test}$ . In other words, the purpose of training the model is just to use the model to classify  $X_{unknown}$ . In general, the main difference between inductive semi-supervised learning and transductive semi-supervised learning is whether the prediction samples encountered during training are the same as the samples to be classified.

Based on the learning scenario, semi-supervised learning methods can also be classified in accordance with the problems, e.g., classification, regression, clustering, and dimensionality reduction<sup>[33,34]</sup>. Most studies on semi-supervised learning have focused on the classification, as most machine learning research does. In particular, in this case, there are typically four types of problems: discriminant learning, generative learning, disagreement-based learning, and semi-supervised graph learning<sup>[35]</sup>. The two typical problems are stated as follows:

1. Generative Learning is an early type of semi-supervised learning that is involved a cyclic process and a



**Figure 4.** Transductive semi-supervised learning diagram.

maximum likelihood parameter estimation procedure. The two main steps are given as follows:

Step 1: Compute the posterior probability of unlabeled data  $P_{\theta} = (C_1|x^u)$ , Initialization is:

$$\theta = \{P(C_1), P(C_2), \mu^1, \mu^2, \sum\} \quad (1)$$

Step 2: Update the algorithm

$$P(C_1) = \frac{N_1 + \sum_{x^u} P(C_1|x^u)}{N} \quad (2)$$

$$\mu^1 = \frac{1}{N_1} \sum_{x^r \in C_1} x^r + \frac{1}{\sum_{x^u} P(C_1|x^u)} \sum_{x^u} P(C_1|x^u) x^u \dots \quad (3)$$

Then, back to step 1.

In equation (1),  $\theta$  means initialization, and the posterior probability of unlabeled data depends on model  $\theta$ . The  $N$  means the total number of examples, and the  $N_1$  means the number of examples belonging to  $C_1$ .

2. Low-density Separation (also known as Self-training) works as follows:

In [Figure 5](#), given the  $X_{labeled} = \{(x^r, \hat{y}^r)\}_{r=1}^R$  and the  $X_{unlabeled} = \{x^u\}_{u=1}^U$ , we could get the  $X_{pseudo} = \{(x^u, y^u)\}_{u=1}^U$  after apply the train model to  $X_{unlabeled}$  as step 2. Steps 1–4 can be continued until the anticipated class labels from step 2 no longer meet a specified probability threshold or until there are no more unlabeled data.

Due to the wide variety of semi-supervised learning methods, each method has a corresponding dataset due to different principles. To evaluate each method fully, there are many evaluation indicators in semi-supervised learning, including accuracy, true positive rate (TPR), false positive rate (FPR), and Receiver Operating Characteristic (ROC) <sup>[37]</sup>, etc. The accuracy is a very common evaluation indicator, so the TPR, FPR, and ROC are detailed below:

1. True positive rate (TPR): is described as the proportion of true positive outcomes among all positive samples.

$$TPR = \frac{TP}{TP + FN} = \frac{Ture Positive}{All Positive Case} \quad (4)$$

A point is said to be True Positive (TP) if this point lies above the Upper Control Limit (UCL) after the damage, which is defined according to a certain damage condition. Conversely, a point is said to be False Negative (FN) if this point lies under the UCL after the damage. Then the UCL signifies the degree of confidence in the training period, as shown in [Figure 6](#).

2. False positive rate (FPR): is described as the proportion of false positive outcomes among all negative samples.

$$FPR = \frac{FP}{FP + TN} = \frac{False Positive}{All Negative Case} \quad (5)$$

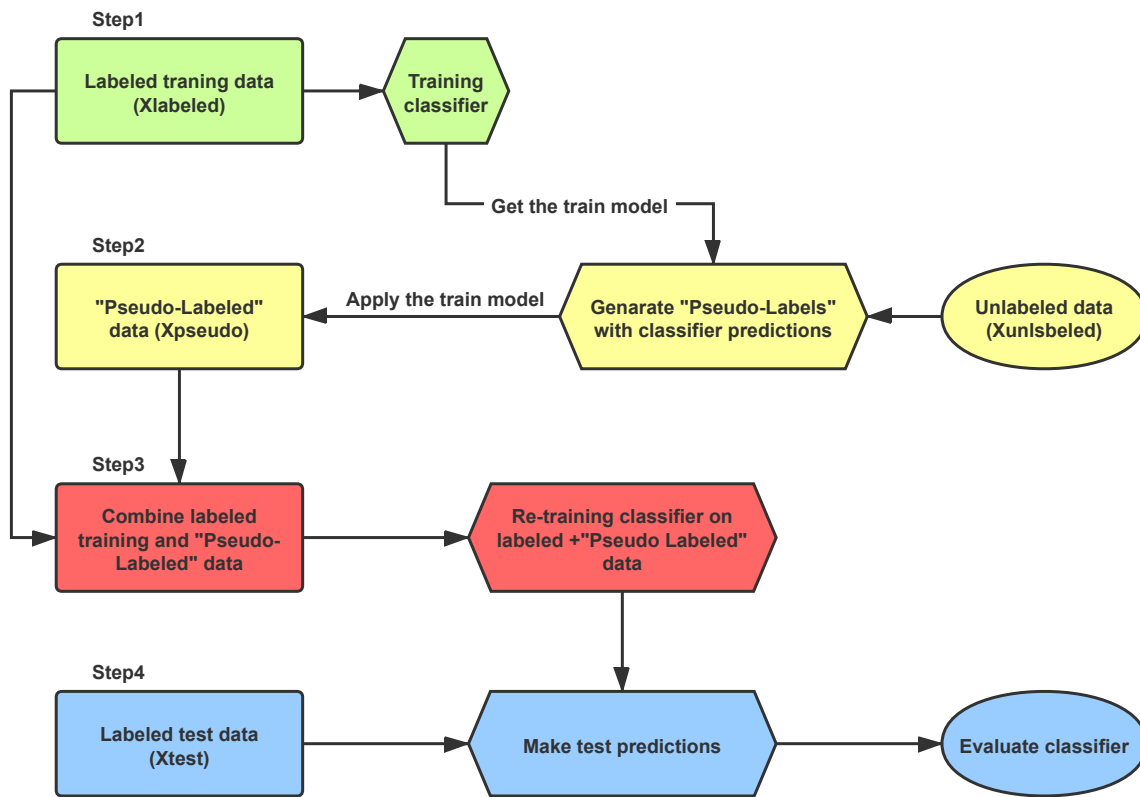


Figure 5. Self-training flowchart [36].

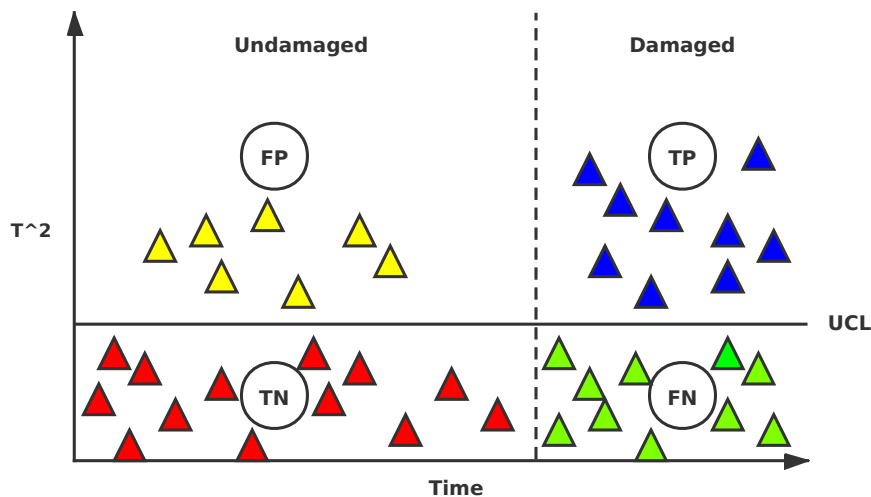
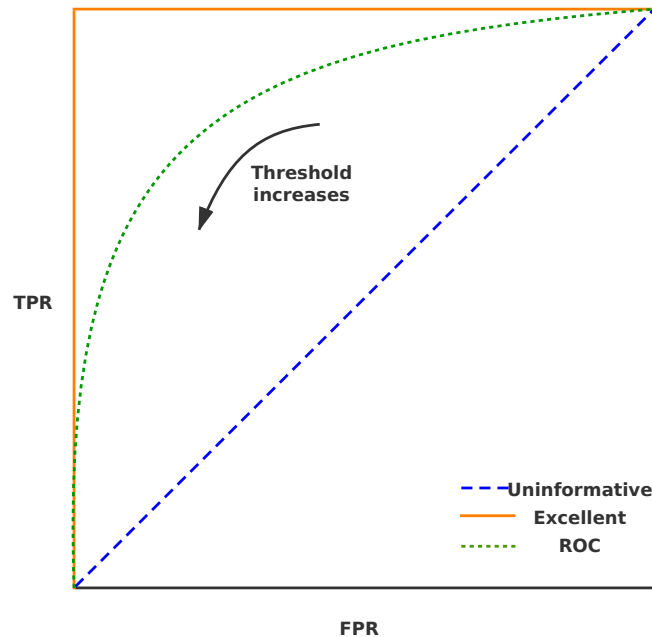


Figure 6. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

A point is defined to be False Positive (FP) if this point lies above the UCL before the damage, while a point is said to be True Negative (TN) if it lies under the UCL.

- Receiver Operating Characteristic (ROC): a ROC curve is plotted by the TPR against FPR at various threshold settings. When the ROC curve is closer to the upper-left-hand corner of the ROC space, the model is





**Figure 7.** Receiver Operating Characteristic (ROC) curve.

more accurate. On the contrary, when the ROC curve is closer to the uninformative line, the model is less accurate as [Figure 7](#).

#### 2.4. Deep learning

Even though the flood is a common phenomenon because it is a natural disaster that occurs in many countries every year, predicting it is rather difficult. Simple machine learning methods and models are not effective in solving complex flood forecasting problems. As a result, deep learning approaches have been centered by both scientists and governments to solve the flood prediction conundrum<sup>[38]</sup>. These approaches help them make accurate predictions and save lives. Roughly speaking, deep learning can be viewed as a subset of machine learning, which is usually employed to process a large amount of data. As such, it makes it possible to train powerful computer programs so as to perform specific tasks, as shown in [Figure 8](#). One of the applications of deep learning can be seen in flood prediction, where a deep learning model can be trained to predict the occurrence of floods using the data collected over many years<sup>[39]</sup>. For this purpose, the National Weather Service (NWS) collects weather data from thousands of local stations across the country. They also collect data from satellite photos aiming to measure the extent of the floods and the damage levels. Using this information, deep models can be created to predict where, how much, and what type of flood conditions will happen subsequently. The results of the prediction are then forwarded to the local governments for pre-emptive planning purposes. This allows authorities to respond quickly to unexpected floods by shutting down streets or opening up new paths through flooded areas.

Although humans are constantly faced with a large amount of perceptual data, they can always obtain important information that is worth noting in a deft way. Imitating the efficiency of the human brain in accurately representing information has long been a core challenge in AI research. Research on the mammalian brain was based on anatomical knowledge: the time when sensory signals traveled from the retina to the prefrontal cortex, which then transferred to the motor nerves, was used to determine how the brain represented information. Inferences were made regarding the cerebral cortex not directly displaying the data but allowing the received stimulus signals to pass through a lamellar network model and then obtaining the observed rules.



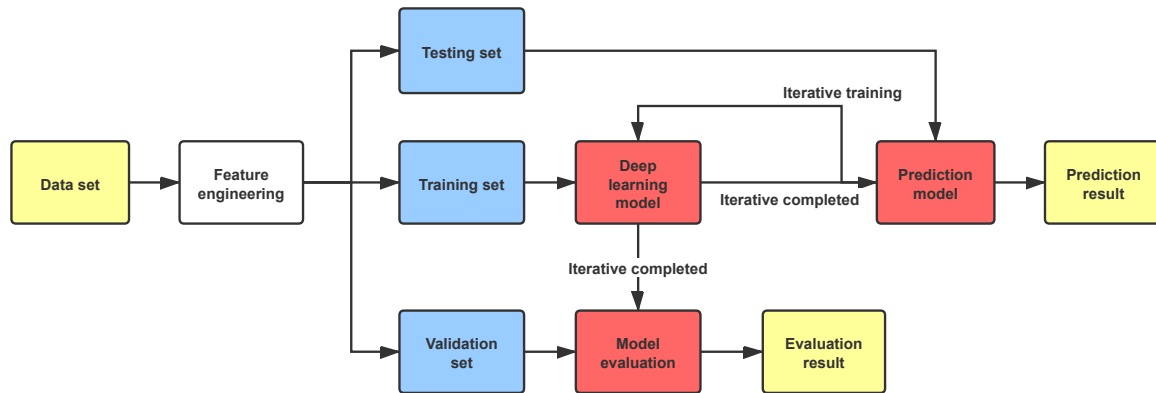


Figure 8. Deep learning block diagram.

Thus, the human brain does not project images directly onto the retina but rather processes information by aggregating and decomposing it to recognize objects. In order for the visual cortex to function properly, it must reproduce the image on the retina and feature the perceptual signal<sup>[40]</sup>. By retaining useful structural information about objects and reducing the amount of data processed by the visual system, this hierarchy of human perception reduces the amount of data that the visual system processes. In addition to capturing essential features in structurally rich data, deep learning can also capture potentially complex structural rules, such as natural images, videos, and voice<sup>[41]</sup>. In the field of machine learning, deep learning is a promising research direction that is aimed at getting machine learning closer to AI. As part of deep learning, the internal rules and representation levels of sample data are learned<sup>[42]</sup>, and the information gained from this process can be used to interpret text, image, and sound data. Deep learning is an algorithm that has achieved far more in speech and image recognition than previous methods related to machine learning. The ultimate goal is for machine learning to be able to recognize words, images, and sounds in a manner similar to humans.

Models for deep learning include convolutional neural networks (CNN), deep reinforcement learning<sup>[43]</sup>, and stacked autoencoder networks<sup>[44,45]</sup>. In the neural and cognitive machine of Fukushima, the first computational model of CNN is proposed, inspired by the visual system<sup>[46,47]</sup>. To obtain a translation invariant neural network structure form, neurons with the same parameters are applied to different positions of the previous layer based on the local connections between neurons and hierarchical image transformation. As a result of this concept, Le Cun *et al.* developed CNN with error gradients<sup>[48]</sup>, which improved pattern recognition performance. Historically, CNN has shown extraordinary performance, such as the handwritten character recognition<sup>[49]</sup>.

The Deep Belief Networks (DBNs) are a type of Bayesian probabilistic generation model<sup>[50,51]</sup>, which can be trained to generate inputs in a probabilistic way. The resulting trained layers can even be used as a feature detector, which is able to be further trained to solve classification problems. For DBNs, a stack of Restricted Boltzmann machines (RBMs) is used, which sometimes consists of two layers of the recurrent neural network (RNN), i.e., the visible layer and hidden layer. Such layers are undirected; thus, an efficient, fast training capability can be obtained when performing the unsupervised procedure. It is observed that due to the network structure of the DBNs, such models can be trained by a greedy learning strategy layer-by-layer. As a result, the DBNs have been found to have a broad range of applications in practice<sup>[52,53]</sup>.

Similar to DBNs, stack autocoding networks consist of several layers of structural units<sup>[54]</sup>. However, the structural units in the autocoding model are auto-encoders instead of RBMs<sup>[55]</sup>. In the autocoding model, there are two layers, i.e., an encoding layer and a decoding layer<sup>[56]</sup>. An effective method for building multi-layer

neural networks based on unsupervised data was proposed by Hinton in 2006<sup>[57]</sup>. A single layer of neurons is constructed in two steps: first, a single layer is built layer by layer, which means a single layer of the network is trained every time; then, the wake-sleep algorithm is used to tune all layers. A bidirectional weighting of all layers except the topmost is used, so the topmost layer remains a single-layer neural network<sup>[58]</sup>, while the other layers become graph models. Use the wake-sleep algorithm to adjust all weights, with the upward weight for "cognition" and the downward weight for "generation". The topmost representation generated must be capable of restoring the underlying node as accurately as possible to ensure cognition and generation agree<sup>[59]</sup>. If a node represents a face at the top, then all face images should activate this node, resulting in an image that can be used to represent a general image of a face.

Currently, to improve the performance of real-time flood forecasting, we could use the Fully Convolutional Network (FCN), Long Short Term Memory model (LSTM), Gated Recurrent Unit Network (GRU), Graph Convolutional Network (GCN), Generative Adversarial Networks (GAN), CNN-Long Short Term Memory Network (CNN-LSTM), and STA-LSTM. Each model or algorithm has a different focus on hydrological prediction and, therefore, has a different effect when we focus on different kinds of flooding.

1. **Fully Convolutional Network (FCN):** A FCN can mainly solve the image segmentation technology from the semantic level. After sampling the feature map of the convolutional layer, the FCN can restore it to the same size as the input image to make pixel predictions. FCN and Multioutput FCN are quite useful in solving problems related to spatial data but do not encode the position and orientation of applied objects. The image-based model of FCN can capture the basic structure of the problem, but it has higher requirements on time and space complexity<sup>[60]</sup>.
2. **Long Short Term Memory (LSTM)/Gated Recurrent Unit Network (GRU):** LSTM networks are designed to have a large short-term memory, allowing for more efficient and resource-rich training on datasets of sequential samples. As a consequence of a series of activation functions and processes, LSTM neurons produce two different values, as compared to a convolution operation generating one output and holding it to urgent receptors in both the next and same layer<sup>[61]</sup>. While both outcomes are kept within the LSTM layer to keep a record of what was learned in the preceding part of the sequential manner, one is transmitted to the next layer<sup>[62]</sup>. GRU and LSTM have comparable performance, and the GRU is the variant of LSTM. However, the parameter of GRU is less, and the training speed is faster than LSTM because the input gate and the forget gate of LSTM are combined with the update gate and the cell state, and the hidden state of LSTM is combined in GRU. The GRU has advantages over LSTM in forecasting when the dataset is huge.
3. **Graph Convolutional Network (GCN):** Graph Neural Network (GNN) refers to the application of neural networks on graphs. According to the classification of propagation mode, GNN can be divided into graph convolutional neural network (GCN) and graph attention network (GAT). GCN can efficiently extract features from non-euclidean structure data, which is prevalent in a large amount of data types lacking regular structure.
4. **Generative Adversarial Networks (GAN):** The GAN introduces the concept of adversarial learning<sup>[63]</sup>. The concept is introduced between the generator and the discriminator. GAN, as presented by Goodfellow *et al.* (2014)<sup>[64]</sup>, presents two separate CNNs that work in unison but are seemingly separate at the same time. It resembles that they are competing against one another in a min-max game. Here, one of the CNN will aim to generate fake examples from the given dataset. Another CNN acts as the discriminator. The role of the discriminator is to test whether the dataset is real or not. Since both the CNNs are working against one another and the training is better, it makes the CNN get better over time. Moreover, Markov models usually compute at a slow speed and might have a high degree of inaccuracy; the GAN can work with high-volume complex data in a faster manner<sup>[65]</sup>.
5. **Convolutional Neural Network LSTM (CNN-LSTM):** The main feature of CNN is the convolution operators. CNN is suitable for processing spatial data because it consists of a convolutional layer and a pooling layer. The convolution layer can maintain the spatial continuity of the image and extract the local features

of the image. The pooling layer can use max-pooling or mean-pooling, reducing the dimension of the middle-hidden layer and the amount of computation. Moreover, LSTM is suitable for processing temporal data. Therefore, CNN-LSTM has better performance than LSTM for spatio-temporal prediction problems.

6. **Spatio-temporal attention LSTM (STA-LSTM):** STA-LSTM includes the main LSTM network and Spatial Attention and Temporal Attention. The main LSTM network is used for feature extraction, spatial-temporal correlation utilization, and final prediction. We can adjust the attention weights dynamically, and the performance of LSTM cells can be improved due to Spatial Attention and Temporal Attention.

Then relevant experimental verification should be conducted appropriately for flood forecasting models. Experimental verification is an essential step in evaluating the performance of a flood forecasting model and ensuring that it is accurate and reliable. The previously unused data (test dataset) can be used to validate the accuracy of the model and use different evaluation metrics to assess its performance. In general, flood forecasting belongs to a regression task, for which evaluation indexes are usually important. Thus, a few commonly used indexes are listed, including Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R Squared, and Nash-Sutcliffe Efficiency (NSE). Relevant experimental verification can and should be conducted appropriately for flood forecasting models. Experimental verification is an essential step in evaluating the performance of a flood forecasting model and ensuring that it is accurate and reliable<sup>[66,67]</sup>.

1. **Mean Squared Error (MSE):** MSE is the expectation of squared error, that is, the average squared deviation between the prediction values and the observation values.

$$MSE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2 \quad (6)$$

In equation (6),  $O_i$  means the observation value,  $P_i$  means the prediction value,  $\bar{O}_i$  means the average observation value and  $n$  means the number of observations. Usually, the MSE could be led to the dimension issues.

2. **Root Mean Square Error (RMSE):** RMSE shows the degree of dispersion of the samples and assesses how well the predicted value matches the observed value. RMSE can avoid the dimension issues compared with MSE. If the machine learning problem is sensitive to the dimension issues, the RMSE could be used as an evaluation index of model performance. Since RMSE is the sum of squared errors and then takes the square root, it would be possible to magnify the gap for the larger errors.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (7)$$

3. **Mean Absolute Error (MAE):** MAE is the expected value of the absolute error loss, that is, the average deviation between the predicted value and actual value. Similar to RMSE, the value of the MAE should be optimized as low as possible too. Since the MAS directly takes the absolute value of the error, the MAS reflects the real error.

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n} \quad (8)$$

4. **R Squared ( $R^2$ ):**  $R^2$  is called the coefficient of determination. The performance of the model cannot be accurately judged when the evaluation index does not have the upper and lower limits, while  $R^2$  is able to take this into account, confining the index value into a certain bound.

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \quad (9)$$

5. **Nash-Sutcliffe Efficiency (NSE):** A normalized statistic, called the NSE, can evaluate the performance of the model to measure variables and can display the percentage of the initial variance of the model.

$$NSE = \left(1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2}\right) \times 100 \quad (10)$$

Note that when models deal with regression problems, if the models have higher performance, the values of the MSE, RMSE, and MAE are as low as possible, but the  $R^2$  and NSE are both nearly 1.

### 3. PROBLEMS AND CHALLENGES OF DIFFERENT ALGORITHMS FOR FLOOD PREDICTION

After analyzing the machine learning working mechanism, this section aims to understand the distinctions and problems between different machine learning methods and propose corresponding models to solve the problems. The researchers must comprehend the characteristics and performance of every machine learning approach while selecting models. Understanding the drawbacks of each algorithm, in particular, can help us better pick models for flood prediction.

#### 3.1. Problems and challenges in supervised learning

When an algorithm or model is proposed, the following aspects need to be considered in order to obtain a good learning performance:

1. **The trade-off between bias and variance:** To generate more precise prediction results, We make a trade-off between bias and variance<sup>[68]</sup>. Generally, bias and variance are a trade-off in a learning algorithm, so a lower bias learning algorithm needs to be flexible so that it can match the data well. However, it is likely that if the learning algorithm is too flexible, it will match every training dataset, which results in a high variance. Therefore, most supervised learning methods offer a bias/variance parameter that the user can adjust so that they can adjust this trade-off between bias and variance. For flood forecasting, the bias (accuracy) would be sacrificed for lower variance (stability) because lower variance means better robustness and generalization ability.
2. **The complexity of the function and the amount of training data:** The second issue is the amount of complexity of the training data relative to the "real" function (classification or regression)<sup>[69]</sup>. With a small amount of data, an "inflexible" learning algorithm can be used to learn a simple function with a high bias and low variance<sup>[70]</sup>. The function, however, will only be trained from a very large number of data, and the use of "flexible" learning algorithms has a low bias and variance if it involves many different input elements in a complex interaction and it has behavior in various parts of the input space<sup>[71]</sup>. Based on data availability and perceived complexity of a function, good learning algorithms automatically adjust bias/variance trade-offs.
3. **The dimension of the input space:** There is also the problem of the input space dimension<sup>[72]</sup>. Even if the true function only depends on a small number of input features, the learning problem is difficult if the input feature vectors have high dimensionality. As a result, high input dimensionality usually requires adjustments to classifiers with low variance and high bias. This is because many of the "extra" sizes can be confounded by the learning algorithm and make it have a high variance. Engineers can improve the accuracy of this learning function by manually removing irrelevant features from the input data. Further, many algorithms are available for choosing features that are relevant and ignoring those that are not.
4. **Output value in noise:** The fourth issue is the level of noise at the desired output values (monitoring the target variables)<sup>[73]</sup>. If the desired output value is usually incorrect (because of human or sensor errors), the learning algorithm should not try to find a training example that exactly matches those values. Trying to fit the data too cautiously leads to overfitting<sup>[74]</sup>. When there is no measurement error (random noise), if you are struggling to learn the function, it is your learning pattern itself that is too complex, potentially leading to overfitting.

#### 3.2. Problems and challenges in unsupervised learning

Unsupervised learning is not commonly employed in real-time flood prediction models<sup>[75]</sup> due to its susceptibility to subjectivity and lack of clear analytical objectives, such as response prediction. Furthermore, there is no widely accepted approach for performing cross-validation or validating the outcomes of unsupervised learning techniques on different datasets, making it difficult to evaluate their effectiveness<sup>[76,77]</sup>.

However, Chen *et al.* (2022) suggested the Flood Domain Adaptation Network (FloodDAN) strategy, which combined adversarial domain adaptation and extensive pretraining to create a model for unsupervised flood forecasting<sup>[78]</sup>. They undertook adversarial domain adaptation between the two datasets after pre-training the source model on large-scale datasets. The final model was built using the source prediction head and target encoder produced in the first and second stages, respectively. The findings of the experiment demonstrated that FloodDAN could make flood predictions using simple rainfall data. In addition, Chen *et al.* (2022) compared the performances of fully supervised learning and unsupervised learning in flood forecasting. In fully supervised learning, researchers used the entire training set for supervised training and compared the evaluation findings to determine the optimum model structure for flood prediction. They computed the lower bound of the unsupervised learning method using the historical runoff input as the model output. The findings demonstrated a typical issue in supervised learning, which was when the amount of data diminished, the performance of the model degraded. However, the FloodDAN unsupervised learning model may perform at the same level as supervised learning which took 450–500 hours, which is very significant and valuable for flood forecasting in regions without hydrological data.

### 3.3. Problems and challenges in semi-supervised learning

While semi-supervised learning has found extensive use in flood prediction, there remain certain challenges in its implementation process. Specifically, two key aspects warrant consideration:

1. **Sample division:** Semi-supervised learning divides samples into unlabeled and labeled samples, which is important for building models<sup>[79]</sup>. In the process of model training by predicting labeled samples and then building a new model, the process is cyclic iteration; and in the process of prediction, it is necessary to consider that the sample prediction is wrong, which will greatly affect its generalization performance<sup>[80]</sup> and even lead to performance degradation. Building a model that can guarantee prediction accuracy without over-labeling is an important challenge in semi-supervised learning<sup>[81]</sup>.
2. **Selection of semi-supervised learning methods:** Semi-supervised learning is not easy to build better learning programs from unlabeled data. As mentioned earlier, unlabeled data are useful if and only if they contain information useful for predicting labels, which do not contain data that cannot be extracted or are difficult to extract in labeled data. Semi-supervised learning methods can be applied in practice to provide effective information for model building, which further leads practitioners and researchers to the question of when such situations occur. It is currently difficult to precisely define the conditions for any particular semi-supervised learning job, and it is not easy to assess the extent to which these conditions are met by the methods that can be used. This question has so far been left unanswered. Research on this problem stops at inferring the applicability of different learning methods to various types of problems<sup>[82]</sup>, such as graph-based methods<sup>[83]</sup> that apply local similarity measures to construct graphs of all data points<sup>[84]</sup>.

### 3.4. Problems and challenges in deep learning

While deep learning methods have decent fitting ability and robustness, there are still some problems that cannot be overlooked.

1. **Expensive computation and low portability:** Deep learning is expensive since it takes a lot of data and processing power. Additionally, a lot of programs are still incompatible with mobile devices. Many businesses and teams are now working on creating semiconductors for portable devices<sup>[85]</sup>.
2. **Strict hardware specifications:** Deep learning demands a lot of computational power, which is something that standard CPUs can no longer handle<sup>[86]</sup>. GPU and TPU are the primary components used in mainstream computing; therefore, both the cost and hardware requirements are relatively high.
3. **Intricate model design:** Deep learning model creation is quite complicated, taking a lot of time, labor, and material resources to create new algorithms and models. Most individuals are limited to using pre-made models. For example, the STA-LSTM and CNN-LSTM can somewhat resolve the robustness and gradient issues, and GRU is able to perform calculations more quickly than LSTM<sup>[87,88]</sup>.



**Table 1. Comparison of machine learning algorithms in the flood forecasting**

Algorithms	Computational cost	Other benefits	Disadvantages
SVM	High on large dataset [91]	Effective in handling high-dimensional data; robust to noise; accurate in nonlinear problems; good for small datasets	The input data needs to be normalized. Sensitivity to kernel choice; parameter tuning difficulty
ANN	High on large dataset	Capture complex nonlinear relationships; better robustness; the ability to learn and adapt to changes in the input data; better generalization	Overfitting easily; parameter tuning difficulty; poor interpretability; sensitivity to input data
CNN	High on large dataset	Spatial and temporal feature extraction ability; No pressure for high-dimensional data processing; Robustness to noise;	Limited applicability; need to adjust parameters and a large and high-quality sample size
FCN	Higher than CNN	handle high-resolution data; learn complex spatial patterns; handle missing data and noise; computationally efficient	Need a large amount and high-quality of labeled data for training; not explicitly model temporal dynamics [60].
GCN	Higher than ANN and CNN	Deal with non-Euclidean data; learn the spatial dependencies between nodes; deal with missing data and noise in the input data; computationally efficient; handle large graphs [92]	Loss of spatial information; requires a well-defined graph structure; not explicitly model temporal dynamics
LSTM	Higher than CNN and ANN	Solved Gradient problem; model temporal dependencies and capture long-term dependencies in time-series data; handle missing data and noise; computationally efficient; handle large datasets [61]	With the increase of time, the modeling ability of LSTM begins to decline due to the gradient problem; overfitting; struggle with handling sudden changes in the data distribution; not explicitly model spatial dependencies [93]
GRU	Lower than LSTM	Alleviate the gradient problem; high memory efficiency; higher computationally efficient; capturing temporal dependencies; good generalization ability [94].	Gradient problem as LSTM; training difficulty; require careful tuning of hyperparameters; sensitivity to initialization; overfitting; limited interpretability
CNN-LSTM	Higher than LSTM	Spatial and temporal feature extraction; improved accuracy; better robustness; better generalization	Overfitting; gradient problem; computational complexity
STA-LSTM	Higher than CNN-LSTM	The high robustness and generalization of hydrological prediction, and the prediction accuracy better than most forecasting models [95]	lower computational efficiency than CNN-LSTM; carefully tune each attention weight; the identification of important spatio-temporal features [11]

#### 4. COMPARISONS AND FURTHER DIRECTION FOR FLOOD FORECASTING

As a result of our determination of the requirements for choosing the right machine learning algorithm, we have listed them below in descending order, starting with the most important and going down to the least important in the prediction of a flood as the following: accuracy, robustness, generalization, low configuration, low preprocessing of data, and insights into factors influencing prediction.

##### 4.1. Characteristics of machine learning algorithms for flood prediction

As many machine learning models have been used for flood prediction, we select some representative models to compare. Table 1 shows the comparison between various machine learning algorithms used in flood prediction. Firstly, SVM has good performance on both low-dimensional and high-dimensional datasets and has certain robustness [89]. Nonetheless, due to the high computational complexity, SVM is usually not efficient for the case when the training datasets are rather large. The input datasets of the flood forecasting model not only have a very large size but also have a relatively high dimension [90]. Therefore, SVM seems to be inadequate to deal with the flood forecasting problem. Moreover, the flood forecasting problem is a nonlinear regression problem; an overfitting issue would more likely occur when the SVM deals with nonlinear problems.

In Artificial Neural Networks (ANN), 3-layer neural network models are employed as the function estimators in order to forecast the flood. Unlike conventional physical models described by certain differential equations, ANN models are able to be trained from the historical data, after which the future trend of the flood can be predicted through the obtained models [96,97]. In this procedure, the actual hydrological datasets are utilized

for training the ANN models; the connecting weights of each neuron are adjusted accordingly in order to fit well the relationship between the occurrence of the flood and related impact factors<sup>[98]</sup>. Clearly, the quality of the provided dataset is one of the most influential features for the ANN performance<sup>[99]</sup>. Due to the fact that the real-time datasets of flood forecasts are sometimes subject to the large noise, data preprocessing is quite necessary in order to improve the ANN models. The data cleaning, normalization, and transformation are good preprocessing techniques that are commonly employed to enhance the quality of the dataset<sup>[90,100]</sup>. Besides that, overfitting is another common issue in ANN model training, particularly when the models used are overly complex or the size of the dataset is relatively small. To this end, the dropout regularization and weight decay are frequently utilized to prevent overfitting, thus improving the generalization ability and robustness<sup>[101]</sup>. Notice that basic ANN architecture merely includes three layers: input layer, hidden layer, and output layer, and usually, such a 3-layer structure cannot achieve the flood forecasting with high accuracy.

The CNN is a kind of ANN architecture trained with deep-learning algorithms<sup>[102]</sup>. CNNs are well suited for solving prediction problems with spatial datasets. This is because CNNs are designed to handle the spatial structure of data and can learn to extract features hierarchically from the input data. The accuracy of a CNN depends on several aspects, such as the size and complexity of the model architecture, the quality of the training data, and the choice of hyperparameters<sup>[103]</sup>. CNNs can be robust in flood forecasting, particularly when trained on a diverse range of data that captures variations in environmental conditions and flood events. This can help the model to identify features that are associated with flooding and to generalize well to new and unseen data. Therefore, CNNs can be made less sensitive to changes in the input data, such as changes in weather conditions or changes in the landscape, which may possibly affect the accuracy of the model. As with most machine learning methods, overfitting may occur as well when the model employed is too complex or the training data is too limited, leading to poor generalization performance on new datasets. Similar to ANN, overfitting would be prevented by the appropriate regularization techniques. Furthermore, the CNN models can be combined with other models to improve the performance of flood forecasting.

The FCN, as a specific type of CNN architecture, has been designed for image segmentation tasks. FCNs and CNNs can be adopted for different types of tasks in flood forecasting due to their different characteristics. The accuracy, robustness, and generalization of FCNs are much similar to CNNs<sup>[104]</sup>. FCNs may sometimes require more computational resources than CNNs owing to their architecture, which could be a limitation for real-time flood forecasting applications. Techniques such as the use of a smaller model architecture or transfer learning can be employed to reduce the computational cost of training an FCN. Mu (2022) trained the models of FCNs, multi-output FCNs, and their RNN variants in watersheds with frequent rainfall and performed both quantitative and qualitative analysis for specific rainfall events<sup>[60]</sup>. In cases where the predicted water depth exceeded 50 cm, the multi-output FCNs had a very significant advantage. Under the recurrent effect, the accuracy of long-term flood prediction of FCNs will be greatly improved. In the process of this test, the three predicted NSE values of the FCN model were 81.0%, 79.95%, and 78.16%, respectively. From this result, this model had a certain accuracy for predicting floods, but the values were not particularly close to 1. The single FCN model here still has a lot of room for optimization. FCNs and CNNs are both useful in flood forecasting, but their specific strengths and weaknesses depend closely on the type of data and task at hand.

GCNs are a class of deep learning models that can effectively handle graph-structured data, making them promising tools for flood forecasting. GCNs have good robustness. For example, they can deal with noisy data and missing values by learning from the local connections in the graph, allowing the model to smooth out noise in the data. GCNs can also effectively adapt to changes in the data by dynamically updating the weights of the connections in the graph, enabling the model to learn from new data without requiring a full retraining<sup>[105]</sup>. Notice that GCNs are mainly designed to handle graph-structured data, which makes them well-suited to handle complex spatial relationships in flood forecasting data. GCNs may struggle to generalize to new and unseen data outside of the training set, especially when the graph structure is significantly different from the



training data. This can lead to overfitting and reduce performance in real-world flood forecasting applications. GCNs may be computationally expensive, especially for large graphs with many nodes and edges. This can make it challenging to scale up GCN-based flood forecasting models in order to cover larger geographic areas. Then Graph convolutional RNN (GCRNN) is a kind of neural network based on GCN. The GCN is used to capture image data to represent spatial relationships, while the RNN is used to capture temporal data<sup>[87]</sup>. The combination of these two structures gives this model an advantage over time and space. GCRNN can be used to predict the time series of water quantity in a geographic area or water supply area. GCRNN is an image-based model that captures water volume in both time and space. Zanfei *et al.* (2022) have tested this prediction model in the presence of sensor failures<sup>[92]</sup>. The test results showed that GCRNN can accurately predict floods when considering the spatial criteria of time series. Especially in fault testing, GCRNN performed much better than LSTM. Since GCRNN has relatively high complexity and the demand for computing time is also high, there are still some deficiencies on the whole. While GCRNN is able to predict the whole flood cycle reliably and stably, the modeling ability begins to decline with the increase of time. From this point of view, GCRNN still has higher advantages. In the model test of the three time periods in the Chongqing section of the Yangtze River Basin, the GCRNN model has significantly higher accuracy than the FCN model since the NSE results of the GCRNN in these three tests are 93.10%, 89.33%, and 83.56%. These three values are larger and closer to 1 than the respective values of FCNs.

LSTM is a type of RNN, and the gradient problem of RNNs can be partially resolved by LSTM models. LSTM models have been shown with high accuracy in flood forecasting compared to traditional statistical models. This is because LSTM models are able to capture the complex temporal dependencies and patterns for the time series data<sup>[106]</sup>. LSTM models have also been shown with the ability to effectively model the nonlinear and non-stationary behavior of hydrological variables, which is important in flood forecasting<sup>[107,108]</sup>. However, the accuracy of LSTM models can be affected by several factors, such as overfitting, underfitting, and the presence of outliers or anomalies in the data. The quality and quantity of input data can significantly impact the robustness of an LSTM model. The architecture of an LSTM model, including the number of layers, hidden units, and input/output dimensions, can impact its robustness too. Moreover, the choice of training parameters, e.g., learning rate, batch size, and optimization algorithms, may also impact the robustness of an LSTM model. The use of appropriate training parameters can improve the ability of the model to generalize to new data. LSTMs performed very well for one-day, two-day, and three-day forecasts. The NSE for these three predictions were 95%, 93%, and 88%, respectively, which are better than FCNs and GCNs.

GRU and LSTM have comparable performance, whereas GRU has a lower computational cost than LSTM. The NSE values of GRU in the three predictions are nearly 95%, 93%, and 88%, respectively. Some researchers combined GRU and CNN models to create the Convolutional GRU (CONV-GRU), aiming to maximize the benefits of both models<sup>[88]</sup>. The combination is accomplished by connecting the output of the CNN model with the input of the GRU model. This model was used to forecast the water level and flood phenomena of a Taiwanese river. This model was thought to be an extension of the GRU model. When compared to other neural network models, the CONV-GRU model was found to be superior to other neural network methods in predicting water level characteristics, and it was very useful in detecting local flood characteristics. The CONV-GRU model can detect both normal and abnormal time series behavior. Furthermore, the error between the predicted and actual value of this model is relatively small, and the CONV-GRU model is superior to the four models of LSTM, CNN, ANN, and Seq2seq (Sequence to sequence).

Although the applications of machine learning techniques in flood prediction have been increasing, most of them are based on a collection of one-dimensional data. The CNN-LSTM model uses the two-dimensional radar map to calculate the runoff and then achieves the purpose of flood prediction<sup>[109]</sup>. The two-dimensional processing of the precipitation radar map by CNN and the one-dimensional processing of the sequence by LSTM makes it possible to figure out the upstream and downstream flow series<sup>[86]</sup>. The database in three

**Table 2. Measures of machine learning algorithms**

Measures	Algorithms
High accuracy	STA-LSTM <sup>[11,93]</sup>
Good robustness	SVM, STA-LSTM
Good generalization	SVM, STA-LSTM, GRU
Low computational cost	SVM, ANN, CNN
Gradient problems	CNN, LSTM, GRU
Fitting problems	SVM, ANN, CNN, FCN, GCN
High dataset quality	SVM, CNN, Hybrid LSTM, GRU

different years was taken as the study period. After forecasting at three different water level periods, the NSE values were 93.51%, 94.25%, and 95.18%, respectively. Comparing this set of results with the NSE values of the LSTM, it was clear that the CNN-LSTM predictions were more accurate. After using the NSE and evaluating the performance of the model, it is found that the better prediction performance of CNN-LSTM highly depends on the optimized input dataset<sup>[110]</sup>. It means while models can be useful for estimating water levels and issuing flood warnings, high-quality datasets are fundamental to their value.

It is observed that most prevailing deep learning structures, such as LSTM networks, are unlikely to model the spatial correlation of hydrological data and, thus, fail to produce satisfactory prediction results<sup>[95]</sup>. It is clear that flood is uncertain and highly nonlinear, which leads to the low robustness of hydrological prediction<sup>[111]</sup>. The physical interpretability of the models is easily overlooked by general machine learning methods. Some scholars proposed the STA-LSTM model based on the original LSTM, which contains an explainable attention mechanism. Experiments showed that this model outperformed FCN, CNN, GCN, and the conventional LSTM in most cases. The rationality of this model is mainly reflected by the visual and interpretable weight of spatial and temporal attention. When carrying out modeling experiments, Lyu *et al.* (2021) found that in the STA-LSTM model, spatio-temporal attention and time have similar weights, and they all move forward slowly. In terms of the overall weight, its change trend is not obvious and relatively slow<sup>[112,113]</sup>. However, in the case of actual flood input, with the extension of the forecast time, the time attention weight also keeps moving forward, and a certain time point will have a synchronous forward speed with the predicted speed. This implies that the time attention weight is very similar to the whole fusion process<sup>[114]</sup>. Moreover, if the weight of time attention deviates, it is highly likely to cause the deviation of the prediction results. From this perspective, although the STA-LSTM model has better performance than other neural networks in terms of datasets and performance, it is greatly affected by the weight of time attention<sup>[115]</sup>. The three values of NSE in this evaluation were 97.03%, 96.73%, and 95.10%, respectively, which shows STA-LSTM had the highest accuracy in predicting floods.

All in all, the high-quality datasets, the great quantity of data, concise model structure, suitable hyperparameters, and data preprocessing are important for the accuracy, robustness, and generalization of flood forecasting models. As a result, the selection of a machine learning algorithm for flood forecasting should be based on the unique requirements of the application and the properties of the available data.

#### 4.2. Future direction for real-time flood forecasting

Although the challenges of flood forecasting have been partially addressed by current machine learning methods, there are still challenges not completely being solved in flood forecasting, as shown in Table 2. The vanishing gradient problem is a common challenge in deep learning-based flood forecasting models<sup>[116]</sup>. LSTM models can relieve the vanishing gradient problem of RNN and CNN models, but it also occurs when the gradient of the loss function becomes very small or very large, making it difficult for the optimizer to update the model parameters effectively during training. Various techniques, such as weight initialization, batch normalization, gradient clipping, and data augmentation, can be used to address the gradient problem and improve the performance of these models<sup>[117]</sup>.

An ideal flood forecasting model should exhibit low bias and low variance to accurately predict true values while not being overly sensitive to minor input data changes<sup>[118]</sup>. However, striking this balance can be challenging, especially when dealing with complex flood systems that are influenced by a multitude of factors and variables. The variance impacts the robustness and generalization of the flood forecasting model, while the bias affects its accuracy<sup>[119]</sup>. Therefore, the balance between variance and bias needs to be adjusted according to the specific forecasting requirements.

A lot of flood forecasting models are sensitive to the input data changing. Because flood events can be affected by various factors, such as weather patterns and land-use changes, future work could focus on developing machine learning models that can adapt to changing conditions and provide accurate predictions under different scenarios. In other words, the flood forecasting models of high generalization stability have to be given more attention. For example, the hybrid CNN models could be created since CNN models are hypersensitive to the input data changing<sup>[120,121]</sup>.

Notwithstanding the fact that machine learning models have flexibility, speed, and simplicity with physical models, their generalization ability is restricted<sup>[122]</sup>. The physical models are based on the underlying physical laws and can generalize well to different conditions and locations. Therefore, hybrid models that combine the strengths of machine learning and physical models can improve flood prediction accuracy and generalization.

The quality and quantity of the dataset will have a large impact on the accuracy of the flood forecasting models. In some regions, data scarcity can limit the effectiveness of deep learning models. Developing methods to address data scarcity is imminent, such as transfer learning and data augmentation. The concept of GAN-LSTM is to generate synthetic flood data that can be used to augment the training data of an LSTM model. GANs can generate realistic and diverse samples from a given distribution, while the LSTM learns to predict the next flood event based on the input data and the synthetic data<sup>[123]</sup>. This approach has the potential to improve the accuracy and generalization of the LSTM models, as it can learn to predict floods based on a wider range of data. Researchers could try to explore the use of advanced deep learning algorithms, such as GANs, to improve the accuracy and robustness of flood forecasting.

## 5. CONCLUSIONS

The purpose of this manuscript is to provide a survey of the current state of machine learning applications in flood prediction. Due to the abundance of literature on the topic, the focus of this review is to analyze and compare the performance of mainstream algorithms that are currently being used in real-time flood prediction.

Accurately detecting the timing and magnitude of major floods is a challenge for watershed managers, as it is critical to provide timely early warnings to those at risk and save lives. Recent advancements in remote sensing technology and the installation of real-time flood water level detection sensors, in conjunction with advanced machine learning techniques, have made it possible to provide more accurate and longer forecast windows for predicting the timing and magnitude of future flooding events. This improved prediction capability enables better countermeasures, evacuation efforts, and mobilization of emergency management teams.

By utilizing advanced remote sensing techniques and real-time flood water level monitoring sensors, a large volume of data can be collected, quickly analyzed, and processed by machine learning algorithms to forecast floods. These algorithms can identify complex patterns in real-time data to accurately predict flood water levels in complex river networks and urban sewer sheds. The data collected can be used to create flood inundation maps and issue warnings to impacted residents through mobile applications on their cell phones, enabling emergency response teams to mobilize and coordinate evacuation efforts more effectively.

In conclusion, this review summarizes the combination of remote sensing, real-time monitoring, and machine learning technology, offering a promising solution to the challenge of accurately forecasting floods and reducing their impact on at-risk communities.

## DECLARATIONS

### Authors' contributions

Made substantial contributions to the research and investigation process, reviewed and summarized the literature, and wrote and edited the original draft: Zhang Y

Conducted the research activity and execution, collaborated on writing the article: Zhang Y, Pan D

Oversight, leadership responsibility, commentary, and critical review: Yang SX, Gharabaghi B, Van Griensven J

### Availability of data and materials

Not applicable.

### Financial support and sponsorship

This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Grant 401643. Grant co-funded by Lakes Environmental Research Inc.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2023.

## REFERENCES

1. Mofitakhari HR, AghaKouchak A, Sanders BF, Allaire M, Matthew RA. What is nuisance flooding? Defining and monitoring an emerging challenge. *Water Resour Res* 2018;54:4218-27. [DOI](#)
2. Zhang M, Conti F, Le Sourne H, et al. A method for the direct assessment of ship collision damage and flooding risk in real conditions. *Ocean Eng* 2021;237:109605. [DOI](#)
3. Luiz-Silva W, Oscar-Júnior AC. Climate extremes related with rainfall in the State of Rio de Janeiro, Brazil: a review of climatological characteristics and recorded trends. *Nat Hazards* 2022;114:713-32. [DOI](#)
4. Bozorg O. Review on IPCC reports. *Climate Change in Sustainable Water Resources Management* 2022:123. [DOI](#)
5. Guo Y, Wu Y, Wen B, et al. Floods in China, COVID-19, and climate change. *The Lancet Planet Health* 2020;4:e443-44. [DOI](#)
6. Yamamoto H, Naka T. Quantitative analysis of the impact of floods on firms' financial conditions. Bank of Japan; 2021.
7. Romero M, Finke J, Rocha C. A top-down supervised learning approach to hierarchical multi-label classification in networks. *Appl Netw Sci* 2022;7:1-17. [DOI](#)
8. Henriksen HJ, Roberts MJ, van der Keur P, et al. Participatory early warning and monitoring systems: a nordic framework for web-based flood risk management. *Int J Disast Risk Re* 2018;31:1295-306. [DOI](#)
9. Ferrans P, Torres MN, Temprano J, Sánchez JPR. Sustainable Urban Drainage System (SUDS) modeling supporting decision-making: a systematic quantitative review. *Sci Total Environ* 2022;806:150447. [DOI](#)
10. Brunner MI, Slater L, Tallaksen LM, Clark M. Challenges in modeling and predicting floods and droughts: a review. *Wiley Interdiscip Rev: Water* 2021;8:e1520. [DOI](#)
11. Ding Y, Zhu Y, Wu Y, Jun F, Cheng Z. Spatio-temporal attention LSTM model for flood forecasting. In: 2019 International Conference on Internet of Things (IThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE; 2019. pp. 458-65. [DOI](#)

12. De La Cruz R, Olfindo Jr N, Felicen M, et al. Near-real-time Flood Detection From Multi-temporal Sentinel Radar Images Using Artificial Intelligence. *ISPRS* 2020;43. DOI
13. Belabid N, Zhao F, Brocca L, Huang Y, Tan Y. Near-real-time flood forecasting based on satellite precipitation products. *Remote Sens* 2019;11:252. DOI
14. Munawar HS, Hammad AW, Waller ST. A review on flood management technologies related to image processing and machine learning. *Autom. Constr* 2021;132:103916. DOI
15. Bronfman NC, Cisternas PC, Repetto PB, Castañeda JV. Natural disaster preparedness in a multi-hazard environment: characterizing the sociodemographic profile of those better (worse) prepared. *PLoS one* 2019;14:e0214249. DOI
16. Chhajer P, Shah M, Kshirsagar A. The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. *Decision Analytics Journal* 2022;2:100015. DOI
17. Zhang J, Bargal SA, Lin Z, et al. Top-down neural attention by excitation backprop. *Int J Comput Vision* 2018;126:1084-102. DOI
18. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Bio* 2022;23:40-55. DOI
19. Rustam F, Reshi AA, Mehmood A, et al. COVID-19 future forecasting using supervised machine learning models. *IEEE access* 2020;8:101489-99. DOI
20. El Boujnouni M. A study and identification of COVID-19 viruses using N-grams with Naïve Bayes, K-nearest neighbors, artificial neural networks, decision tree and support vector machine. In: 2022 International Conference on Intelligent Systems and Computer Vision (ISCV). IEEE; 2022. pp. 1-7. DOI
21. Cunningham P, Cord M, Delany SJ. Supervised learning machine learning techniques for multimedia. Springer; 2008. DOI
22. Seel NM. Encyclopedia of the sciences of learning. Springer Science & Business Media; 2011. DOI
23. Zhou ZH. A brief introduction to weakly supervised learning. *Natl Sci Rev* 2018;5:44-53. DOI
24. de Bruijn JA, de Moel H, Jongman B, et al. A global database of historic and real-time flood events based on social media. *Sci data* 2019;6:311. DOI
25. Khan W, Ghazanfar MA, Azam MA, et al. Stock market prediction using machine learning classifiers and social media, news. *J Amb Intel Hum Comp* 2020:1-24. DOI
26. Van Engelen JE, Hoos HH. A survey on semi-supervised learning. *Mach Learn* 2020;109:373-440. DOI
27. Ghahramani Z. Unsupervised learning. Advanced Lectures on Machine Learning. LNAI 3176. Springer-Verlag; 2004. DOI
28. Zhou ZH, Zhou ZH. Semi-supervised learning. *Mach Learn* 2021:315-41. DOI
29. Sammut C, Webb GI. Encyclopedia of machine learning. Springer Science & Business Media; 2011. DOI
30. Jukes E. Encyclopedia of machine learning and data mining. *Reference Reviews* 2018;32:3-4. DOI
31. Zhu XJ. *Semi-supervised learning literature survey* 2005.
32. Mey A, Loog M. Improved generalization in semi-supervised learning: a survey of theoretical results. *IEEE T Pattern Anal* 2022. DOI
33. Xu W, Tang J, Xia H. A review of semi-supervised learning for industrial process regression modeling. In: 2021 40th Chinese Control Conference (CCC). IEEE; 2021. pp. 1359-64. DOI
34. Yang X, Song Z, King I, Xu Z. A survey on deep semi-supervised learning. *IEEE T Knowl Data En* 2022:1-20. DOI
35. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiat* 2020;77:534-40. DOI
36. Gull T, Khurana S, Kumar M. Semi-supervised labeling: a proposed methodology for labeling the twitter datasets. *Multimed Tools Appl* 2022 03;81. DOI
37. Giglioni V, García-Macías E, Venanzi I, Ierimonti L, Ubertini F. The use of receiver operating characteristic curves and precision-versus-recall curves as performance metrics in unsupervised structural damage classification under changing environment. *Eng Struct* 2021;246:113029. DOI
38. Opella JMA, Hernandez AA. Developing a flood risk assessment using support vector machine and convolutional neural network: a conceptual framework. In: 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA). IEEE; 2019. pp. 260-65. DOI
39. Sankaranarayanan S, Prabhakar M, Satish S, et al. Flood prediction based on weather parameters using deep learning. *J Water Clim Change* 2020;11:1766-83. DOI
40. Arzoumanian Z, Baker PT, Blumer H, et al. The NANOGrav 12.5 yr data set: search for an isotropic stochastic gravitational-wave background. *The Astrophysical journal letters* 2020;905:L34. DOI
41. Benetos E, Dixon S, Duan Z, Ewert S. Automatic music transcription: an overview. *IEEE Signal Proc Mag* 2018;36:20-30. DOI
42. Zhou T, Thung KH, Zhu X, Shen D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Hum Brain Mapp* 2019;40:1001-16. DOI
43. Lin J, Li J, Chen J. An analysis of English classroom behavior by intelligent image recognition in IoT. *Int J Syst Assur Eng* 2021:1-9. DOI
44. Chen S, Yu J, Wang S. One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes. *J Process Contr* 2020;87:54-67. DOI
45. Masarczyk W, Głomb P, Grabowski B, Ostaszewski M. Effective training of deep convolutional neural networks for hyperspectral image classification through artificial labeling. *Remote Sens* 2020;12:2653. DOI
46. Lindsay GW. Convolutional neural networks as a model of the visual system: Past, present, and future. *J Cognitive Neurosci* 2021;33:2017-31. DOI
47. Zeman AA, Ritchie JB, Bracci S, Op de Beeck H. Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex. *Sci Rep-Uk* 2020;10:2453. DOI



48. Alzubaidi L, Fadhel MA, Olewi SR, Al-Shamma O, Zhang J. DFU\_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. *Multimed Tools Appl* 2020;79:15655-77. DOI
49. Li Z, Wu Q, Xiao Y, Jin M, Lu H. Deep matching network for handwritten Chinese character recognition. *Pattern Recogn* 2020;107:107471. DOI
50. Devaraj J, Madurai Elavarasan R, Shafiullah G, Jamal T, Khan I. A holistic review on energy forecasting using big data and deep learning models. *Int J Energ Res* 2021;45:13489-530. DOI
51. Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: Proceedings of the 24th international conference on Machine learning; 2007. pp. 473-80. DOI
52. Imamverdiyev Y, Abdullayeva F. Deep learning method for denial of service attack detection based on restricted boltzmann machine. *Big data* 2018;6:159-69. DOI
53. Satarzadeh E, Sarraf A, Hajikandi H, Sadeghian MS. Flood hazard mapping in western Iran: assessment of deep learning vis-à-vis machine learning models. *Nat Hazards* 2022;1-19. DOI
54. Ying C, Li Q, Liu J. A Brief Investigation for Techniques of Deep Learning Model in Smart Grid. In: 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM). IEEE; 2021. pp. 173-81. DOI
55. Raza K, Singh NK. A tour of unsupervised deep learning for medical image analysis. *Curr Med Imaging* 2021;17:1059-77. DOI
56. Amin SU, Alsulaiman M, Muhammad G, Mekhtiche MA, Hossain MS. Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Gener Comp Sy* 2019;101:542-54. DOI
57. Zhang Y, Wu J, Cai Z, Du B, Philip SY. An unsupervised parameter learning model for RVFL neural network. *Neural Networks* 2019;112:85-97. DOI
58. Mittal S, Lamb A, Goyal A, et al. Learning to combine top-down and bottom-up signals in recurrent neural networks with attention over modules. In: International Conference on Machine Learning. PMLR; 2020. pp. 6972-86.
59. Forbus KD, Ferguson RW, Lovett A, Gentner D. Extending SME to handle large-scale cognitive modeling. *Cognitive Sci* 2017;41:1152-201. DOI
60. Mu Y. An evaluation of deep learning models for urban floods forecasting; 2022.
61. Sit M, Demiray BZ, Xiang Z, et al. A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci Technol* 2020;82:2635-70. DOI
62. Nevo S, Morin E, Gerzi Rosenthal A, et al. Flood forecasting with machine learning models in an operational framework. *Hydrol Earth Syst Sc* 2022;26:4013-32. DOI
63. Saxena D, Cao J. Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)* 2021;54:1-42. DOI
64. Goodfellow IJ. On distinguishability criteria for estimating generative models. *arXiv preprint arXiv:14126515* 2014. DOI
65. Salazar A, Vergara L, Safont G. Generative adversarial networks and markov random fields for oversampling very small training sets. *Expert Syst Appl* 2021;163:113819. DOI
66. Le XH, Ho HV, Lee G, Jung S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* 2019;11:1387. DOI
67. Ren Q, Li M, Song L, Liu H. An optimized combination prediction model for concrete dam deformation considering quantitative evaluation and hysteresis correction. *Adv Eng Inform* 2020;46:101154. DOI
68. Yang Z, Yu Y, You C, Steinhardt J, Ma Y. Rethinking bias-variance trade-off for generalization of neural networks. In: International Conference on Machine Learning. PMLR; 2020. pp. 10767-77.
69. Wang Q, Ma Y, Zhao K, Tian Y. A comprehensive survey of loss functions in machine learning. *Annals of Data Science* 2020;1-26. DOI
70. Anisha P, Polati A. A bird eye view on the usage of artificial intelligence. In: Communication Software and Networks: Proceedings of INDIA 2019. Springer; 2021. pp. 61-77. DOI
71. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *P Natl Acad Sci* 2019;116:15849-54. DOI
72. Heinlein A, Klawonn A, Lanser M, Weber J. Combining machine learning and adaptive coarse spaces—a hybrid approach for robust FETI-DP methods in three dimensions. *SIAM J Sci Comput* 2021;43:S816-38. DOI
73. Jiang Y, Yin S, Dong J, Kaynak O. A review on soft sensors for monitoring, control, and optimization of industrial processes. *IEEE Sens J* 2020;21:12868-81. DOI
74. Arnott R, Harvey CR, Markowitz H. A backtesting protocol in the era of machine learning. *The Journal of Financial Data Science* 2019;1:64-74. DOI
75. Sulaiman J, Wahab SH. Heavy rainfall forecasting model using artificial neural network for flood prone area. In: IT Convergence and Security 2017: Volume 1. Springer; 2018. pp. 68-76. DOI
76. Chen D, Liu F, Zhang Z, Lu X, Li Z. Significant wave height prediction based on wavelet graph neural network. In: 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAl). IEEE; 2021. pp. 80-85. DOI
77. MacKinnon DP. Introduction to statistical mediation analysis. Routledge; 2012. DOI
78. Chen D, Zhou R, Pan Y, Liu F. A simple baseline for adversarial domain adaptation-based unsupervised flood forecasting. *arXiv preprint arXiv:220608105* 2022. DOI
79. Li J, Socher R, Hoi SC. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:200207394* 2020. DOI
80. Scheinost D, Noble S, Horien C, et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *NeuroImage* 2019;193:35-45. DOI

81. Lowrance CJ, Lauf AP. An active and incremental learning framework for the online prediction of link quality in robot networks. *Eng Appl Artif Intel* 2019;77:197-211. [DOI](#)
82. Cai S, Wang Z, Lu L, Zaki TA, Karniadakis GE. DeepM&Mnet: inferring the electroconvection multiphysics fields based on operator approximation by neural networks. *J Comput Phys* 2021;436:110296. [DOI](#)
83. Guo Q, Zhuang F, Qin C, et al. A survey on knowledge graph-based recommender systems. *IEEE T Knowl Data En* 2020;34:3549–68. [DOI](#)
84. Kang Z, Pan H, Hoi SC, Xu Z. Robust graph learning from noisy data. *IEEE T Cybernetics* 2019;50:1833-43. [DOI](#)
85. Lyu L, Fang M, Wang N, Wu J. Water level prediction model based on GCN and LSTM. In: 2021 7th International Conference on Computer and Communications (ICCC). IEEE; 2021. pp. 1600-1605. [DOI](#)
86. Yang W, Chen L, Chen X, Chen H. Sub-daily precipitation-streamflow modelling of the karst-dominated basin using an improved grid-based distributed Xinanjiang hydrological model. *J Hydrol-Reg Stud* 2022;42:101125. [DOI](#)
87. Feng J, Wang Z, Wu Y, Xi Y. Spatial and temporal aware graph convolutional network for flood forecasting. In: 2021 International Joint Conference on Neural Networks (IJCNN). IEEE; 2021. pp. 1-8. [DOI](#)
88. Miao S, Hung WH. River flooding forecasting and anomaly detection based on deep learning. *IEEE Access* 2020;8:198384-402. [DOI](#)
89. Taşar B, Kaya YZ, Varçin H, Üneş F, Demirci M. Forecasting of suspended sediment in rivers using artificial neural networks approach. *International Journal of Advanced Engineering Research and Science* 2017;4:237333. [DOI](#)
90. Sahoo A, Samantaray S, Ghose DK. Prediction of flood in Barak River using hybrid machine learning approaches: a case study. *J Geol Soc India* 2021;97:186-98. [DOI](#)
91. Shilton A, Palaniswami M, Ralph D, Tsoi AC. Incremental training of support vector machines. *IEEE T Neural Networ* 2005;16:114-31. [DOI](#)
92. Zanfei A, Brentan BM, Menapace A, Righetti M, Herrera M. Graph convolutional recurrent neural networks for water demand forecasting. *Water Resour Res* 2022;58:e2022WR032299. [DOI](#)
93. Zhang Y, Gu Z, Thé JVG, Yang SX, Gharabaghi B. The discharge forecasting of multiple monitoring station for Humber River by hybrid LSTM models. *Water* 2022;14:1794. [DOI](#)
94. Cho M, Kim C, Jung K, Jung H. Water level prediction model applying a long short-term memory (lstm)-gated recurrent unit (gru) method for flood prediction. *Water* 2022;14:2221. [DOI](#)
95. Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z. Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* 2020;403:348-59. [DOI](#)
96. Dtiisibe FY, Ari AAA, Titouna C, Thiare O, Gueroui AM. Flood forecasting based on an artificial neural network scheme. *Nat Hazards* 2020;104:1211-37. [DOI](#)
97. Ahmad M, Al Mehedi MA, Yazdan MMS, Kumar R. Development of machine learning flood model using Artificial Neural Network (ANN) at Var River. *Liquids* 2022;2:147-60. [DOI](#)
98. Hassanpour Kashani M, Montaseri M, Lotfollahi Yaghin MA. Flood estimation at ungauged sites using a new hybrid model. *J Appl Sci* 2008;8:1744-49. [DOI](#)
99. Tabbussum R, Dar AQ. Comparative analysis of neural network training algorithms for the flood forecast modelling of an alluvial Himalayan river. *J Flood Risk Manag* 2020;13:e12656. [DOI](#)
100. Jabbari A, Bae DH. Application of Artificial Neural Networks for accuracy enhancements of real-Time flood forecasting in the Imjin Basin. *Water* 2018;10. [DOI](#)
101. Dong P, Liao X, Chen Z, Chu H. An improved method for predicting CO<sub>2</sub> minimum miscibility pressure based on artificial neural network. *Advances in Geo-Energy Research* 2019;3:355-64. [DOI](#)
102. Kimura N, Yoshinaga I, Sekijima K, Azechi I, Baba D. Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions. *Water* 2019;12:96. [DOI](#)
103. Zhang L, Huang Z, Liu W, Guo Z, Zhang Z. Weather radar echo prediction method based on convolution neural network and long short-term memory networks for sustainable e-agriculture. *J Clean Prod* 2021;298:126776. [DOI](#)
104. Sun W, Wang R. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. *IEEE Geosci Remote S* 2018;15:474-78. [DOI](#)
105. Yuan F, Xu Y, Li Q, Mostafavi A. Spatio-temporal graph convolutional networks for road network inundation status prediction during urban flooding. *Comput Environ Urban* 2022;97:101870. [DOI](#)
106. Mehedi MAA, Khosravi M, Yazdan MMS, Shabanian H. Exploring temporal dynamics of river discharge using univariate Long Short-Term Memory (LSTM) Recurrent Neural Network at east branch of Delaware River. *Hydrology* 2022;9:202. [DOI](#)
107. Liu M, Huang Y, Li Z, et al. The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. *Water* 2020;12:440. [DOI](#)
108. Song T, Ding W, Wu J, et al. Flash flood forecasting based on long short-term memory networks. *Water* 2019;12:109. [DOI](#)
109. Li X, Xu W, Ren M, Jiang Y, Fu G. Hybrid CNN-LSTM models for river flow prediction. *Water Supply* 2022;22:4902–19. [DOI](#)
110. Li P, Zhang J, Krebs P. Prediction of flow based on a CNN-LSTM combined deep learning approach. *Water* 2022;14:993. [DOI](#)
111. Kasiviswanathan KS, He J, Sudheer K, Tay JH. Potential application of wavelet neural network ensemble to forecast streamflow for flood management. *J Hydrol* 2016;536:161-73. [DOI](#)
112. Lin L, Li W, Bi H, Qin L. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms. *IEEE Intel Transp Sy* 2021;14:197-208. [DOI](#)
113. Noor F, Haq S, Rakib M, et al. Water level forecasting using spatiotemporal attention-based Long Short-Term Memory Network. *Water*



- 2022;14:612. [DOI](#)
114. Wang Y, Huang Y, Xiao M, et al. Medium-long-term prediction of water level based on an improved spatio-temporal attention mechanism for long short-term memory networks. *J Hydrol* 2023;618:129163. [DOI](#)
  115. Chen C, Luan D, Zhao S, et al. Flood discharge prediction based on remote-sensed spatiotemporal features fusion and graph attention. *Remote Sens* 2021;13:5023. [DOI](#)
  116. Liu M, Chen L, Du X, Jin L, Shang M. Activated gradients for deep neural networks. *IEEE T Neur Net Lear* 2021. [DOI](#)
  117. Luo Y, Huang Z, Zhang Z, Wang Z, Li J, et al. Curiosity-driven reinforcement learning for diverse visual paragraph generation. In: Proceedings of the 27th ACM International Conference on Multimedia; 2019. pp. 2341-50. [DOI](#)
  118. Chang FJ, Hsu K, Chang LC. Flood forecasting using machine learning methods. MDPI; 2019. [DOI](#)
  119. Tran DA, Tsujimura M, Ha NT, et al. Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam. *Ecol Indic* 2021;127:107790. [DOI](#)
  120. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. pp. 427-36. [DOI](#)
  121. Hendrycks D, Gimpel K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:161002136* 2016. [DOI](#)
  122. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* 2021;64:107-15. [DOI](#)
  123. Cheng M, Fang F, Navon I, Pain C. A real-time flow forecasting with deep convolutional generative adversarial network: Application to flooding event in Denmark. *Phys Fluids* 2021;33:056602. [DOI](#)