CHAPTER 2

LITERATURE SURVEY

[1] Developing mathematical models using Artificial Neural Networks

Artificial Neural Networks (ANNs) are efficient mathematical modeling systems with efficient parallel processing, enabling them to mimic the biological neural network using inter-connected neuron units. Among all ML methods, ANNs are the most popular learning algorithms, known to be versatile and efficient in modeling complex flood processes with a high fault tolerance and accurate approximation. In comparison to traditional statistical models, the ANN approach was used for prediction with greater accuracy. ANN algorithms are the most popular for modeling flood prediction since their first usage in the 1990s. Instead of a catchment's physical characteristics, ANNs derive meaning from historical data. Thus, ANNs are considered as reliable data-driven tools for constructing black-box models of complex and nonlinear relationships of rainfall and flood, as well as river flow and discharge forecasting. ANNs were already successfully used for numerous flood prediction applications, e.g., stream-flow forecasting, river flow, rainfall–runoff, precipitation–runoff modeling, water quality, evaporation, river stage prediction, low-flow estimation, river flows, and river time series. Despite the advantages of ANNs, there are a number drawbacks associated with using ANNs in flood modeling, e.g., network architecture, data handling, and physical interpretation of the modeled system. A major drawback when using ANNs is the relatively low accuracy, the urge to iterate parameter tuning, and the slow response to gradient-based learning processes. Further drawbacks associated with ANNs include precipitation prediction and peak-value prediction.

[2] Training the network using Multilayer Perceptron (MLP)

The vast majority of ANN models for flood prediction are often trained with a back-propagation neural network. While back-propagation neural networks are widely used today in this realm, the MLP—an advanced representation of ANNs— recently gained popularity. The MLP is a class of feed-forward neural network which utilizes the supervised learning of BP for training the network of interconnected nodes of multiple layers. Simplicity, nonlinear activation, and a high number of layers are characteristics of the MLP. Due to these characteristics, the model was widely used in flood prediction and other complex hydrogeological models. In an assessment of ANN classes used in flood modeling, MLP models were reported to be more efficient with better generalization ability. Nevertheless, the MLP is generally found to be more difficult to optimize. Back-percolation learning algorithms are used to individually calculate the propagation error in hidden network nodes for a more advanced modeling approach.

[3] Predictive modeling using Decision Trees (DT)

The machine learning method of DT is one of the contributors in predictive modeling with a wide application in flood simulation. DT uses a tree of decisions from branches to the target values of leaves. In classification trees, the final variables in a DT contain a discrete set of values where leaves represent class labels and branches represent conjunctions of features labels. When the target variable in a DT has continuous values and an ensemble of trees is involved, it is called a regression tree. Regression and classification trees share some similarities and differences. As DTs are classified as fast algorithms, they became very popular in ensemble forms to model and predict floods. The classification and regression tree, which is a popular type of DT used in ML, was successfully applied to flood modeling. The random forests method is another popular DT method for flood prediction. Random forests include a number of tree predictors. Each individual tree creates a set of response predictor values associated with a set of independent values. Another major DT is the M5 decision-tree algorithm. M5 constructs a DT by splitting the decision space and single attributes, thereby decreasing the variance of the final variable. Further DT algorithms popular in flood prediction include reduced-error pruning trees, Naïve Bayes trees, chi-squared automatic interaction detectors, logistic model trees, alternating decision trees etc.

[4] Extracting Information from Sources with Wavelet Neural Networks

Wavelet transform (WT) is a mathematical tool which can be used to extract information from various data sources by analyzing local variations in time series. WT has significantly positive effects on modeling performance. Wavelet transforms supports the reliable decomposition of an original time series to improve data quality. The accuracy of prediction is improved through discrete WT (DWT), which decomposes the original data into bands, leading to an improvement of flood prediction lead times. DWT decomposes the initial data set into individual resolution levels for extracting better-quality data for model building. DWTs, due to their beneficial characteristics, are widely used in flood time-series prediction.

In flood modeling, DWTs were widely applied in, e.g., rainfall–runoff, daily stream-flow, and reservoir inflow. Furthermore, hybrid models of DWTs, e.g., wavelet-based neural networks, which combine WT, feed-forward neural network, and wavelet-based regression models, which integrate WT and multiple linear regression, were used in time-series predictions of floods.

[5] Adaptive Neuro - Fuzzy Inference System (ANFIS)

The fuzzy logic is a qualitative modeling scheme with a soft computing technique using natural language. Fuzzy logic is a simplified mathematical model, which works on incorporating expert knowledge into a fuzzy inference system (FIS). An FIS further mimics human learning through an approximation function with less complexity, which provides great potential for nonlinear modeling of extreme hydrological events, particularly floods. Adaptive neuro-FIS, or so-called ANFIS, is a more advanced form of neuro-fuzzy based on the T–S FIS, first coined. Today, ANFIS is known to be one of the most reliable estimators for complex systems. ANFIS technology, through combining ANN and fuzzy logic, provides higher capability for learning. This hybrid ML method corresponds to a set of advanced fuzzy rules suitable for modeling flood nonlinear functions.

[6] Reducing Expected Errors with the help of Support Vector Machine

Hearst Et Al proposed and classified the support vector (SV) as a nonlinear search algorithm using statistical learning theory. Later, the Support Vector Machine (SVM) was introduced as a class of SV, used to minimize over-fitting and reduce the expected error of learning machines. SVM is greatly popular in flood modeling; it is a supervised learning machine which works based on the statistical learning theory and the structural risk minimization rule. The training algorithm of SVM builds models that assign new non-probabilistic binary linear classifiers, which minimize the empirical classification error and maximize the geometric margin via inverse problem solving. SVM is used to predict a quantity forward in time based on training from past data.

[7] Ensemble Prediction Systems (EPSs)

A multitude of ML modeling options was introduced for flood modeling with a strong background. Thus, there is an emerging strategy to shift from a single model of prediction to an ensemble of models suitable for a specific application, cost, and dataset. ML ensembles consist of a finite set of alternative models, which typically allow more flexibility than the alternatives. Ensemble ML methods have a long tradition in flood prediction. In recent years, ensemble prediction systems (EPSs) were proposed as efficient prediction systems to provide an ensemble of N forecasts. In EPS, N is the number of independent realizations of a model probability distribution. EPS models generally use multiple ML algorithms to provide higher performance using an automated assessment and weighting system. Such a weighting procedure is carried out to accelerate the performance evaluation process. The advantage of EPS is the timely and automated management and performance evaluation of the ensemble algorithms. Therefore, the performance of EPS, for flood modeling in particular, can be improved. EPSs may use multiple fast-learning or statistical algorithms as classifier ensembles, e.g., ANNs, MLP, DTs, rotation forest (RF) bootstrap, and boosting, allowing higher accuracy and robustness.