IOP Conf. Series: Earth and Environmental Science 1091 (2022) 012040 doi:10.1088/1755-1315/1091/1/012040

Review of flood prediction hybrid machine learning models using datasets

Ainaa Hanis Zuhairi[1], Fitri Yakub[1], Sheikh Ahmad Zaki[1], Mohamed Sukri Mat Ali[1]

[1] iKohza Wind Engineering for (Urban, Artificial, Man-made) Environment Laboratory, Level 10, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia.

*Corresponding author e-mail: ainaahanis@graduate.utm.my, mfitri.kl@utm.my

Abstract. Floods are among the most destructive natural disasters, and they are extremely difficult to model. Over the last two decades, machine learning (ML) methods have made significant contributions to the advancement of prediction systems that provide better performance and cost-effective solutions by mimicking the complex mathematical expressions of physical flood processes. Because of the numerous benefits and potential of ML, its popularity has skyrocketed. Researchers hope to discover more accurate and efficient prediction models by introducing novel ML methods and hybridising existing ones. The main focus of this paper is to show the state of the art of hybridising ML models in flood prediction. The most effective strategies for improving ML methods are hybridization, data decomposition, algorithm ensemble, and model optimization.

1. Introduction

Extreme weather occurrences have become more common in many regions of the world, possibly as a result of a shift in the climatic scenario [1]. Identification of natural catastrophes in earlier stage like floods, can considerably aid us in limiting the degree of damage caused by these events. This leads to requirement for efficient flood management system, effective and accurate forecasting system is one of the most important elements of any flood management programme. Machine learning is a problem-solving strategy based on bionics that replicates the operation of the human brain, it is often composed of a sequence of mathematical operations and transformations carried out by computer programmes linked to current electronic processors. It has the potential of becoming a system that is economical and dependable in flood prediction.

2. ML Methods in Flood Prediction

Historical flood event records, as well as cumulative data in real time from variety of sensors as well as rain gauges over varied intervals, are frequently utilised to create the ML prediction model. The dataset's conventional sources are usually rainfall as well as water level, which are monitored with rain gauges on land or by more recent remote sensing methods using satellites technology, multi sensor systems, and or using radars [2].

2.1 Short Term and Long-Term Flood Forecasting Methods

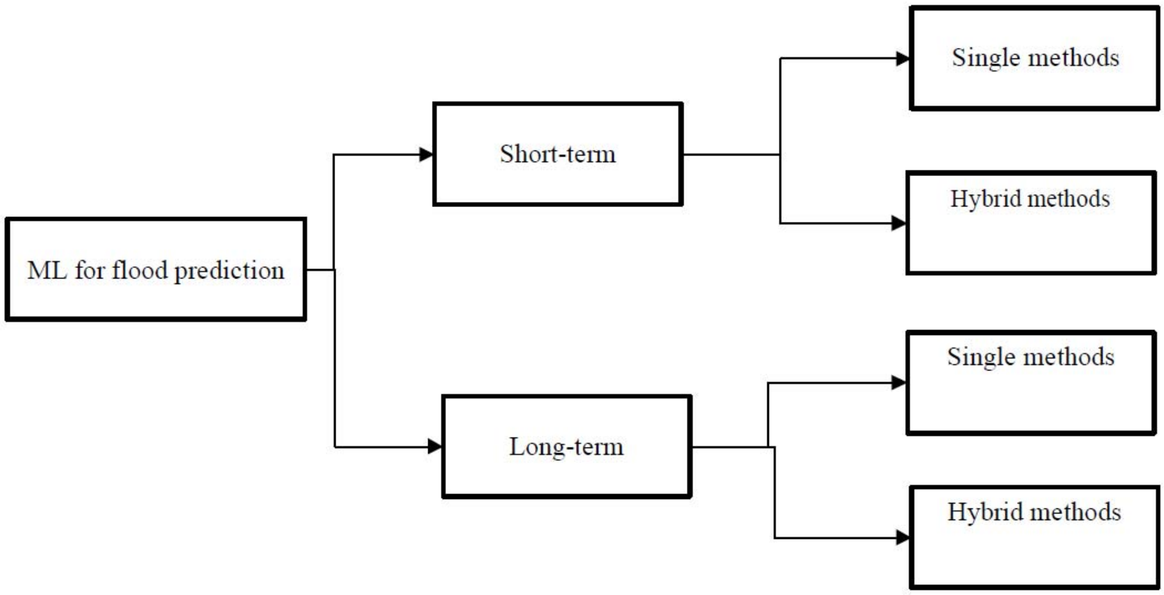

Figure 1. Machine Learning prediction methods [3].

ML has demonstrated considerable success in developing models that identifies patten in fields ranging from computer vision to recognising speech, understanding text and game Artificial Intelligence or AI [4 - 7]. Aside from these classical or traditional fields, ML, particularly deep learning, is becoming increasingly relevant and successful in engineering as well as sciences [5 – 7]. These success stories are founded on the data-driven nature of the approach to learning deriving from a massive number of cases. A recent review synthesises this into a new paradigm of theory-guided data science, emphasising the need of scientific consistency in machine learning [8].

There is even a survey regarding the incorporation of knowledge into support vector machines [9]. The combination of symbolic and connectionist AI appears to be becoming more accessible. We allude to a recent assessment on graph neural networks and a research approach defined as relational inductive bias [10] in this regard. In this study machine learning with hybrid models are reviewed. There are two types of methods for flood prediction approach, short-term and long-term. Short-term flood forecasts are regarded as major research or study issue, especially within densely populated zone, in order to provide early warnings to residents and therefore limit damage [11]. Furthermore, short-term forecasts assist significantly towards water resource management sector.

In spite of recent advancements in terms of numbers, weather prediction or Numerical Weather Prediction (NWP) models, artificial intelligence (AI) and ML, short-term forecasting is still a difficult endeavour [12 -18]. Building hybrid ML approaches to increase prediction quality with regards to accuracy, generalisation, longer lead time, uncertainty, computing costs, and speed is becoming increasingly popular. There are several hybrid approaches, with more well-known models like ANFIS, which is Adaptive Network based

Fuzzy Inference model as well as adaptive neuro fuzzy inference model and WNN or Wavelet Neural Network model, along with other new algorithms.

Long-term flood forecasting is critical towards enhancing knowledge and managing water resource capability across prolonged time periods, ranging from weekly, monthly to yearly forecasts [18]. Many noteworthy ML approaches, such as ANFIS, SVR or Support Vector Regression model, SVM also known as Support Vector Machine model, ANN which is Artificial Neural Network model, bootstrap–ANN and WNN [18 - 22], have been employed for long lead-time predictions with promising results in the previous decades. Recently, the results of several ML approaches for long-term flood forecasts were placed in comparison with a number of publications [21 - 23]. However, it is still unclear whether machine learning algorithm performs the best in long-term flood prediction. A single statistical or machine learning technique is used for classification in an individual approach. Because they incorporate the complimentary advantages of more than one learning approach and overcome the weaknesses of individual techniques, hybrid and ensemble models are efficient and resilient. Stand-alone hybrid models, transformational hybrid models, strongly linked hybrid models, and completely coupled hybrid models are all possible [24].

IOP Conf. Series: Earth and Environmental Science 1091 (2022) 012040 doi:10.1088/1755-1315/1091/1/012040

Table 1. A Review of Flood Prediction Hybrid Machine Learning Models

Author	Title	Models	Objective	Function	Assumption	Accuracy	Conclusion
Li & Willems, 2020	A Hybrid Model for Fast and Probabilistic Urban Pluvial Flood Prediction	Logistic regression and multivariate characterisation	To create a one-of-a-kind hybrid model addressing the issue in regards to models that are data-driven.	It is ideal for real-time flood forecasting in urban catchments due to its capability in providing rapid, generally accurate flood predictions at specified inner-city locations.	Upstream instantaneous inflow depth, Upstream catchment rainfall return period, Downstream outfall rainfall return period, Current node depth, Upstream instantaneous rainfall returns period, and Downstream outfall rainfall return period	Flood alerts with a probability of 50% can have an accuracy of up to 86 %, with a computation saving of 96 %.	The created hybrid model incorporates advantage of both physics-based as well as data-driven models. The suggested hybrid model's computation performance is actually slower when compared to models that use ANN data-driven approaches. Extending the model's boundaries using intensity of rainfall episodes for calibration may result in worse accuracy and less detail than the 1-D HD model.
Kurian et al., 2020	Effective flood forecasting at higher lead times through hybrid modelling framework	M4 hydrological model with ANN	The research suggests a hybrid models of ANN combined with a hydrological model based on physical distribution.	Using observed streamflow as input for prediction accuracy while shortening forecast lead times.	Rainfall, observed streamflow, observed rainfall, forecasted rainfall, simulated streamflow, and simulated streamflow.	Model M4, a 2% variance of noise, with efficiency of 0.95, and 0.94 for a 5% variance of noise.	When paired with assumed flows from HEC-HMS or SWAT with longer lead periods, the forecast accuracy of ANN models rose considerably.

IOP Conf. Series: Earth and Environmental Science 1091 (2022) 012040 doi:10.1088/1755-1315/1091/1/012040

Kan et al., 2020	Hybrid machine learning hydrological model for flood forecast purpose.	ANN and K-nearest neighbour.	continually predicting discharge without diminishing accuracy	Hybridisation increases ANN performance by finding a compromise between network complexity and training accuracy by overcoming the network topology and parameters concurrent global optimization network problem with complicated topology.	Rainfall and antecedent runoff	Most flood occurrences have an NSCE greater than or equal to 0.9.	The suggested model outperforms standard neural network models in terms of continuous discharge prediction without accuracy loss due to its carefully built model structure. In real-world applications, the performance and dependability of the HML hydrological model were shown.
Xie et al., 2021	Hybrid Deep Learning Modelling for Water Level Prediction in Yangtze River	Wavelet transform and LSTM network.	This research uses hybridised deep learning technique LSTM system and discontinuous wavelet transform for daily water level prediction.	Towards better understanding of temporal properties, for the wavelet transform is used to decompose time series into detail approximate part and a novel LSTM model is used with a greedy layer wise unsupervised learning algorithm to learn	Water Level	When compared to other models, the Water Level WA-LSTM model consistently has the lowest MAE and RMSE, which are 46 %, 51 %, 39 %, 50 %, 74 %, and 72 %.	Comprehensive study demonstrates that the maximum performance is obtained by using 5 to 6 days observations lag as features and the Meyer wavelet transform with decomposition level 4., confirming the WA-LSTM model's adaptability and generalisation.

				features of water level via layer-by-layer granulation.			
Yoo et al., 2020	Data Driven Water Surface Elevation Forecasting Model with Hybrid Activation Function—A Case Study for Hangang River, South Korea	Hybrid Activation Function with LSTM	A hybrid model in the form of a blend between hyperbolic tangent function with rectified linear unit using a weighting factor towards water level and long-term forecasting period accuracy improvement.	LSTM in TensorFlow along with open-source software library on Google's deep-learning, was used to predict water level rise in Hangang River, Korea, from 2009 to 2018.	Elevation, Precipitation, Wind Speed, Outlet Discharge, Water Tidal Level and Dissolved Oxygen Concentration.	The predicting accuracy of available data was given as 0.31 m via RMSE while it is 98.9 % via NSE.	The hybrid framework or model presented in this work proved suitable for forecasting increase in water levels in the Hangang River's stated elevation despite a lower forecasting accuracy at the longer leading time interval and a less noticeable improvement in model performance for virtual flood forecasting.
Moishin et al., 2021	Designing Deep-Based Learning Flood Forecast Model With ConvLST	CNN with LSTM	To construct and assess a forecasting model that predicts the occurrence of flood occurrences in the future, this	The study employs a Flood Index (IF) as a mathematical representation, which is used by flood monitoring system to anticipate the next	Precipitation, Antecedent daily rainfall data	All predicted horizons had d values and RNSE that are more than 0.93, 0.85as well as 0.95.	The forecasting approach may not be suitable in places where the applicability of IF has not yet been demonstrated, as the predictive model built only employed two input variables.

	M Hybrid Algorithm		study proposes a combination of deep learning method that combines the predictive advantages of CNN and LSTM Network.	daily IF value, this forecasting model uses IF lag that is statistically significant, enhanced using antecedent and rainfall statistic in real time			
Imran & Sheikh Abdul Khader, 2020	Forecasting Water Level of Jhelum River of Kashmir Valley, Using Prediction and Early Warning System	Water level prediction using ANFIS model and Linear Multiple Regression or LMR model.	Purpose of this study is to suggest alternative framework that is easy to install, low-cost and accurate. The system's LMR method is quick to fit, easy to read, and suitable for forecasting continuous responses like water level.	The framework is made up of a command-and-control centre including base station. The base station is made up of a sensor along with a processing module that performs a localised water level forecasting and transmits projected and observed data to control centre	30 years of daily precipitation and temperature data, as well as daily water-level data from 1980 to 2019.	This hybrid approach has a daily forecast accuracy of 93.53 percent and an hourly prediction accuracy of 99.91 percent.	According to the model, the predicted as well as measured values were nearly similar., with residuals lying between 1, mild inflation variance of a coefficient, the system's accuracy in short-term forecasts is higher than its accuracy in long-term predictions.

3. Discussion and Recommendation

Current status of ML models in regard to flood prediction is still quite young and in its infancy. The purpose of this study is to present a high-level overview of hybrid machine learning system or framework used in flood prediction. The suggested hybrid models have an advantage over standard models in that some can forecast constantly without losing accuracy due to its uniquely built model structure. Though there are limitations, such as, calibration of the models requires comprehensive flood data at localities where it sometimes may be difficult to collect. However, with a growing emphasis for data that are crowdsourced and publicly available gathering as well as citizen research, a more complete flood record data will become more widely available in the future. Despite the promising outcomes of adopting machine learning approaches that have already been published, there has been substantial study and testing for future enhancement and advancement, four key trends for increasing prediction quality were documented. Hybridization, whether through the combination of either two or more ML techniques, the combination of a ML methods with more traditional approaches, with or without soft computing. Data decomposition techniques are used for the goal of increasing dataset quality greatly contributed to boosting prediction accuracy. The usage of an ensemble of techniques, which significantly enhanced the generalisation capacity of the models and reduced prediction uncertainty. The final method was to increase the quality of machine learning algorithms by using optimiser algorithms.

4. Conclusion

Certainly, the growth of these unique ML approaches is heavily reliant on the effective use of soft computing techniques in the construction of fresh learning algorithms and soft computing approaches are key contributors to constructing future hybrid ML systems. These approaches listed in this paper are only but some of the available options for hybridised machine learning system. There are still plenty more though there may have been some reviews and comparison done for hydrological and Neural Network hybrid models, others are still not that well documented or reviewed. This may be another opportunity for future study.

5. References

[1] Lal P N, Mitchell T, Aldunce P, Auld H, Mechler R, Miyan A, Romano L E, Zakaria. S 2012 *National systems for managing the risks from climate extremes and disasters Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (*Special Report of the Intergovernmental Panel on Climate Change* Vol. 9781107025) ed A Dlugolecki and T Masumoto et al (Cambridge University Press, Cambridge and New York) p 339–392.

[2] Seo D J and Breidenbach J P 2002 Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J. Hydrometeorology*. **3** 93

[3] Mosavi A, Ozturk P, Chau K W 2018 *Flood prediction using machine learning models: Literature review* (Water vol 10) ed Chang F. J and Hsu K et al (MDPI: Switzerland) p 1-14

[4] Silver D, Huang A, Maddison C J, Guez A, Sifre L, Van Den Driessche G, Schrittweiese J, Antonoglou L, Panneershelvam V et al. 2016 Mastering the game of Go with deep neural networks and tree search. *J. Nature.* **529** 484.

[5] Kutz J N 2017 Deep learning in fluid dynamics. *J. Fluid. Mechanics*. **814** 1.

[6] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow P M, Zietz M, Hoffman M M et al. 2018 Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface.* **15** 1

[7] Butler K T, Davies D W, Cartwright H, Isayev O, Walsh A 2018 Machine learning for molecular and materials science. *J. Nature* **559** 547

[8] Karpatne A, Atluri G, Faghmous J, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N and Kumar V 2016 Theory-guided Data Science: A New Paradigm for Scientific Discovery from

Data. *J. IEEE. Trans. Knowl. Data. Eng*. **29** 2318
[9] Lauer F and Bloch G. 2008 Incorporating prior knowledge in support vector machines for classification: A review. *J. Neurocomputing*. **71** 1578
[10] Battaglia P W, Hamrick J B, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R et al. 2018 Relational inductive biases, deep learning, and graph networks. *J. arXiv.* v **3** 1
[11] Zhang J, Hou G, Ma B, Hua W 2018 Operating characteristic information extraction of flood discharge structure based on complete ensemble empirical mode decomposition with adaptive noise and permutation entropy. *J. Vib. Control.* **24** 5291
[12] Noori R, Karbassi A, Farokhnia A, Dehghani M 2009 Predicting the Longitudinal Dispersion Coefficient Using Support Vector Machine and Adaptive Neuro-Fuzzy Inference System Techniques. *J. Environ. Eng. Sci*. **26** 1503
[13] Panda R K, Pramanik N, Bala B 2010 Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model. *J. Comput. Geosci*. **36** 735
[14] Kourgialas N N, Dokou Z, Karatzas G P 2015 Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *J. Environ. Manage*. **154** 86
[15] Saghafian B, Haghnegahdar A, Dehghani M 2017 Effect of ENSO on annual maximum floods and volume over threshold in the southwestern region of Iran. *J. Hydrol. Sci.* **62** 1039
[16] Kim G, Barros A P 2001 Quantitative flood forecasting using multisensor data and neural networks. *J. Hydrol*. **246** 45
[17] Badrzadeh H, Sarukkalige R, Jayawardena A 2015 Hourly runoff forecasting for flood risk management: Application of various computational intelligence models. *J. Hydrol*. **529** 1633
[18] Choubin B, KHalighi Sigaroodi S H, Malekian A 2016 Impacts of Large-Scale Climate Signals on Seasonal Rainfall in the Maharlu - Bakhtegan Watershed. *J. Range. Watershed. Management.* **69** 51
[19] Liang Z, Li Y, Hu Y, Li B, Wang J A 2018 Data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. *J. Theor. Appl. Climatol.* **133** 137
[20] Elsafi S H 2014 Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *J. Alexandria. Eng.* **53** 655
[21] Teng J, Jakeman A J, Vaze J, Croke B F W, Dutta D, Kim S 2017 Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environ. Model. Softw.* **90** 201
[22] Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A 2019 An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *J. Sci. Total. Environ*. **651** 2087
[23] Han S, Coulibaly P 2017 Bayesian flood forecasting methods: A review. *J Hydrol*. **551** 340
[24] Bahrammirzaee A, Ghatari A R, Ahmadi P, Madani K 2009 Hybrid credit ranking intelligent system using expert system and artificial neural networks. *J. Appl. Intell.* **34** 28
[25] Li X, Willems P A 2020 Hybrid Model for Fast and Probabilistic Urban Pluvial Flood Prediction. *J. Water. Resour. Res*. **56** 1
[26] Kurian C, Sudheer K P, Vema V K, Sahoo D 2020 Effective flood forecasting at higher lead times through hybrid modelling framework. *J. Hydrol.* **587** 124945.
[27] Kan G, Liang K, Yu H, Sun B, Ding L, Li J, He X, Shen C 2020 Hybrid machine learning hydrological model for flood forecast purpose. *J. Open. Geosci.* **12** 813
[28] Xie Z, Liu Q, Cao Y 2021 Hybrid deep learning modeling for water level prediction in yangtze river. *J. Intell Autom. Soft. Comput.* **28** 153
[29] Yoo H J, Kim D H, Kwon H H, Lee S O 2020 Data driven water surface elevation forecasting model with hybrid activation function-a case study for hangang river, South Korea. *J. Appl. Sci*. **10** 1424

[30] Moishin M, Deo R C, Prasad R, Raj N, Abdulla S 2021 Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. *J. IEEE. Access.* **9** 50982

[31] Imran M, Sheikh Abdul Khader P 2020 Forecasting water level of Jhelum River of Kashmir Valley India, using prediction and earlywarning system. *J. Geogr. Environ. Sustain.* **13** 35