FLOOD PREDICTION MODEL IN MALAYSIA: A REVIEW PAPER

Wei Ming Wong[1], Siva Kumar Subramani[2], Farah Shahnaz Feroz[3], Lew Ai Fen Rose[4]

Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Malaysia (UTeM)[1,4]

Advance Sensors & Embedded Controls System (ASECS), Centre for Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia[2]

Centre for Advanced Computing Technology (C-ACT), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia[3]

Abstract. *Flood is among the deadliest natural disaster in many countries, including Malaysia. Annually, flood happens in Malaysia at two different states, which is either in the form of a flash flood or seasonal flood. One way of understanding or forecasting incoming floods is by designing a reliable flood prediction model with an extended forecast period. Existing of the flood prediction model, the emergency response team has sufficient time to respond. This main paper contribution is to present a state-of-the-art flood prediction model. Researchers in Malaysia have been studying four significant forms of flood prediction models in recent years, which will be addressed in this paper too. These models are the Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), machine learning, and Nonlinear Autoregressive Exogenous Artificial Neural Network (NARX). The accuracy and efficiency of the flood prediction model are essential, and these few factors need to be considered, Root Mean Square Error (RMSE), model best fit, and R-squared (R^2). This paper thus proposes the most ambitious model of flood prediction to be used in Malaysia. This study can be used as a guideline to choose the proper flood prediction model for predicting floods.*

Keywords: *flood, forecast, prediction, Malaysia*

1. INTRODUCTION

A flood happens around the globe, and for Malaysia, it is a crucial problem. It is a natural disaster that becomes an annual event in Malaysia. Floods can no longer be seen and viewed as an isolated event, as they are closely linked to problems such as disease outbreaks, food insecurity, and deterioration of climate. Flood still happens in Malaysia, although the government had executed multiple plans with a large budget, such as the enlargement of drainage[1].

Hundreds of lives have been lost due to floods in the past decades, directly or indirectly. Besides, flood presents the most widespread natural threat to life today, as compared to all the other natural hazards. Based on the 2014 & 2015 data from the Water Resources Management and Hydrology Division, there are a total of 381 flood cases occurred in Malaysia[2]. It is a common situation where the government buildings or schools to be converted as the Evacuation Centre since these locations can shelter many people or flood victims at the same time. These places also need to have essential basic such as drinking water supply. But some of these shelters cause people to feel trapped, hence it is crucial to have the geographic information to let the authority know which is the best location for evacuation and as Evacuation Centre[3].

It is important to forecast flood because of continuous heavy rainfall that caused flood is challenging to avoid in countries that receive high rainfall rates annually, especially in countries like Malaysia. Flood forecasting also became a discussion among world researchers to get the best prediction of flood occurrence.

Flood forecasting is among the most complex to model, but this research needed for risk reduction due to flood. Flood not only damage properties, but it also may take the lives of people. Flood forecasting has become increasingly important. It involves high accuracy and statistical modeling, such as trend analysis. Those modelings are essential to help the local authority to manage and identify the flood trends and patterns as well as the capacity of the flood to harming the current safety.

On 1st December 2019 at Melaka, Malaysia, continuous 2 hours heavy rainfall caused the flash flood, and 20 vehicles where sunk in the water. At the same time, another 30 cars stuck in the car park buildings and unable to move out while waiting for the water to recede[4]. Heavily rain, which falls since 5.30 pm, caused the water to overflow in areas of Lebuh Tun Razak, Bukit Katil, Batu Berendam, and Ayer Keroh. Fig. 1 shows the vehicle sunk in the flood in this disaster.

Figure 1 Vehicle Sunk in the Flood

While on 7th July 2019 at Melaka, Malaysia, a flash flood caused 1096 peoples to be displaced. The flash flood occurred in Alor Gajah and Melaka Tengah caused by the run-off rainwater from storms in Negeri Sembilan, which then overflowed downstream in the state of Melaka. State Disaster Management Committee secretariat head Effendy Ali said as of 5:00 pm, and there were 1021 victims from 229 families in Alor Gajah and another 75 from 18 families in Melaka Tengah[5]. Fig. 2 show the victims evacuate to the evacuation centre after their home had flooded.

Figure 2 Victims at Evacuation Centers

In Malaysia, flash floods and seasonal floods often occur. Flash floods typically happen in the high population area, such as Kuala Lumpur or Melaka. Meanwhile, the seasonal flood occurs when there is a monsoon and have a specified period for the flood to occur. A seasonal flood happens according to a particular time and month. Seasonal floods mainly occur in the southern part of Peninsular Malaysia, including the areas of Melaka and Johor, from December until January due to Northeast Monsoon[3].

The occurrence of flash floods has a strong relationship with rainfall and geography. Rainfall with high intensity in a short period and complex orography tend to have a high frequency of flash flood[6]. Flash flood also occurred due to unplanned building development, which could lead to a poor drainage problem in the city. When there is heavy rain, rainwater could not adequately drain out of the town, which then causes a flash flood. Besides, drastic land use planning and development at river basins could increase the impermeable of land surrounding the river basin, thus increases the volume and peak discharge of hydrograph generated by the river basin[7].

This paper aims to identify various forms of the model for flood prediction. Each type of flood prediction model has different prediction accuracy, and this exactitude also affects the prediction duration of the incoming flood. By discussing the different kinds of flood prediction models, opinions and suggestions to choose the best model will then given to suit and serve the main purpose of predicting flood.

2. FLOOD PREDICTION MODEL

It is essential to have a flood prediction model to forecast the occurrence of floods. wThe flood prediction model needs to be as accurate as possible to provide precise information for the rescue team to respond to the incoming disaster. There is four main flood predictions model discussed in this paper which are ARIMA, SARIMA, machine learning, and NARX.

2.1. ARIMA Time Series

The Autoregressive Integrated Moving Average (ARIMA) was introduced in 1976 by Box and Jenkins. It has since become the most popular model for forecasting data from the univariate time series. The root of the model is from the Moving Average (MA) model and Autoregressive (AR) model. Both of these models, AR and MA, also have a combination called ARMA model. There are three stages in the Box-Jenkins step to construct an ARIMA model, model identification, model estimation, and model checking. These steps are essential for the time-series data to identify the best ARIMA model[8]. Equ. 1 shows the equation of the ARIMA model with the parameter of ARIMA (p, d, q).

AR (p) = Auto regression order of model;
I (d) = Degree of differencing
MA (q) = Moving average of the model

The equation to define the above notation is:

$$y_t = \sum_{i=1}^{p} \emptyset_i \, y_{t-i} + \sum_{j=1}^{p} \theta_j \, e_{t-j} + \epsilon_t + C \qquad (1)$$

$\emptyset$ = Autoregressive model's parameters,
θ = Moving average model's parameters.
c = Constant,
ε = Error terms (white noise).

Where ϵ_t is the error term at time t, $\emptyset_i$ for the variance stabilizing, i represent the parameter number of i-th autoregressive and θ_j represent the parameter number the j-th moving average. ARIMA models used in various applications ranging from medical, economic, and engineering. The application of ARIMA include in the cases such as stock market prediction, electric power consumption, COVID-19 cases, or various of data with sufficient measurement that can represent by time series to be used to be model[9]

The differencing approach will use in the Box–Jenkins methodology to make the original data stationary. The autocorrelation function (ACF) and partial autocorrelation function (PACF) parameter used to decide the term for autoregressive and moving average variable, which to apply in the ARIMA model.[10]. According to [11], the ARIMA model is the most accurate method compare with other prediction techniques in [12].

2.2. SARIMA

Seasonal ARIMA (SARIMA) model is where seasonal components included in the ARIMA model. A combination of the seasonal term in the ARIMA model formed the seasonal ARIMA. Fig. 3 shows an example of the SARIMA equation.

Based on the equation, parameter m = number of observations per year. In this equation, the lowercase notation is for the non-seasonal parts of the model, and uppercase notation used for the seasonal components of the model[13].

$$\text{ARIMA} \quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

Figure 3 Combination of ARIMA with Seasonal Part

2.3. Machine Learning

Machine learning is to give the ability for the system to learn and improve from the previous experience without using the programme by the application of Artificial Intelligence (AI). During the past two decades, there are lots of complex mathematical calculations processes of floods. Machine learning (ML) methods give a high contribution to the improvement of flood prediction systems by providing cost-effective and high performance. For that purpose, the individual data sets are being trained, validated, verified, and tested on basic flow to construct a machine learning model described in Fig. 4.

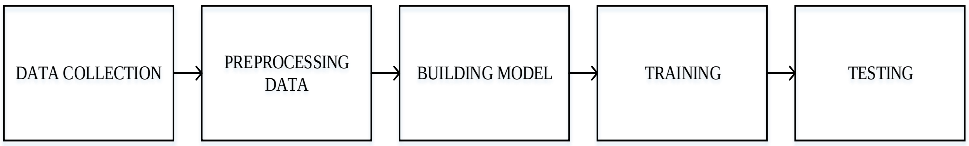

Figure 4 Flow for Building the Machine Learning

2.4. NARX

The Nonlinear Autoregressive Exogenous Artificial Neural Network (NARX) model is a type of dynamically driven recurrent artificial neural network (ANN). Recurrent networks have one or more loops of feedback which can be local or global. Global loops reduce the requirements of computational memory. There are two necessary uses for the recurrent network, which are associative memory and input-output mapping networks. Two applications of input-output are prediction in the form of time series and signal modeling.

The most apparent advantage of NARX models is those different models made up of the same structure, and thus have a fair cost of computation. Therefore, whenever a NARX network uses a period forecast as an input for subsequent periods compared to a feedforward network, degrees of freedom will be achieved. These will allow summary information on exogenous variables to be included as well as a smaller number of residuals, thereby reducing the number of parameters to be estimated[9].

3. METHOD

In this study, the methods and models used for predicting flood prediction identified through the review of journal articles in a similar subject field.

These studies had few variable parameters to consider, which are the location, flood resource variable, time interval of data, root mean square error (RMSE), R^2 or best-fit percentage, type of floods, prediction duration, data obtained, and software used.

The search query for this review includes two main search terms. The first main term of the search is the flood prediction model (flood prediction model). The second main term consists of the four search terms that often used in flood prediction, such as “prediction,” “forecast,” “Malaysia,” and “Estimation.”

Table 1 present the summary of the flood prediction model that had been studied and compiled on the table. The parameter considered in this study are the location, flood resource variable, RMSE, and MSE, R squared and best fit, model types, types of the flood, and prediction duration.

Table 1 Flood Prediction Models

Types of Forecast	Location	Flood Resource Variable	RMSE and MSE	R^2 and Best Fit	Model types	Types of Flood	Prediction Duration
SARIMA [14]	Dungun, Terengganu, Malaysia	Water Level	MSE 0.01246		(0,1,1) (0,1,1)12	Seasonal	Monthly
ARIMA [15]	Segamat River	Rainfall		0.9895	(0,1,2)		
ANN[14]	Dungun, Terengganu, Malaysia	Water Level	MSE 0.00674		4-6-1	Seasonal	Monthly
BPNN (Backpropagation Neural Network)[16]	Dungun, Terengganu, Malaysia	Water Level	MSE 0.0016		5-8-1	Seasonal	Monthly
NARX (Neural Network Autoregressive with Exogenous Input)[16]	Dungun, Terengganu, Malaysia	Water Level	MSE 0.0008		NARX 5-10-1 d= 4	Seasonal	Monthly
NARX[17]	Sungai Besut, Terengganu Sungai Dunggun	Water Level	0.0220	85.47		Flash	5 hours
NNARX (Neural Network Autoregressive with Exogenous Input) [18]	Muda River, Kedah	Water Level	0.0067	95.29	[4,10,1]	Flash	7 hours
ENN (Elman Neural Network) [19]	Kedah	Water Level	0.008	93.6249	[4,10,1]	Flash	
NARX[19]	Kedah	Rainfall, Water Level	0.0203	85.7319	[4,10,1]	Flash	
NNARX[20]	Klang River	Water Level	0.017			Flash	
Multiple-Input Single-Output (MISO) Autoregressive with Exogenous Input (ARX)[21]	Pahang River, Temerloh, Pahang	Water Level	0.01869	50.41	ARX 441	Flash	7 hours

MISO Autoregressive Moving Average with Exogenous Input (ARMAX) [22]	Pahang River, Temerloh, Pahang	Water Level	0.00947	63.06	RMX 2221	Flash	7 hours
Non-linear Auto Regressive Exogenous Neural Network (NARXNN) [22]	Kelantan	Water Level	0.0842	85.56	NARX-NN	Flash	7 hours

4. DISCUSSION

There are different types of flood prediction models with different purpose and accuracy. It also depends on how long the prediction duration. The longer the prediction duration, the percentage of best fit also will decrease, and this will reduce the flood prediction model accuracy. The most accurate model to use for flood prediction is by using machine learning such as NARX. This finding matches the finding from [22].

When deciding on the model to be used, it is crucial to know the purpose of the flood prediction model. There are two purposes in the flood prediction model, which is to predict the flash flood and to predict seasonal flash flood. Multiple types of flood resource variables can use for flood forecastings, such as rainfall, streamflow, and water level.

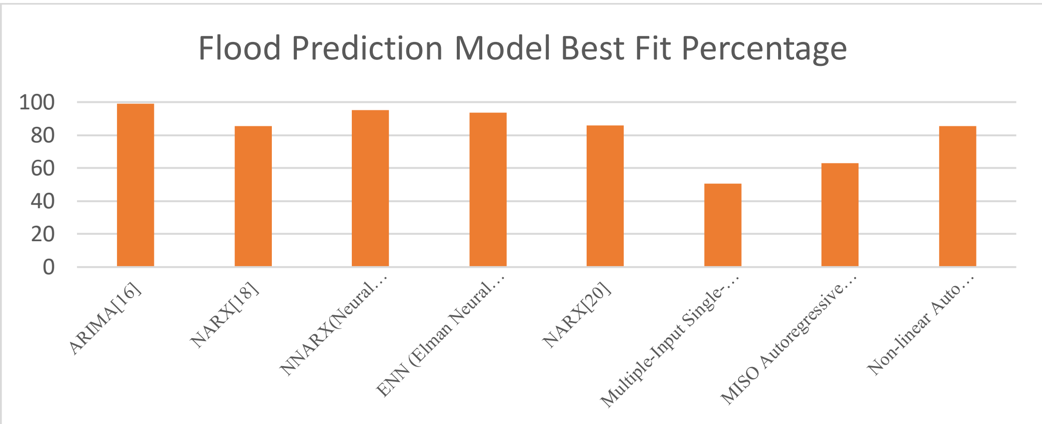

Figure 5 Flood Prediction Model Best Fit Percentage

Fig 5 and Fig 6 shows the result of flood prediction model accuracy based on the model best fit, Mean Square Error (MSE), and Root Mean Square Error (RMSE). The model with a high percentage of best fit and R^2 means that a particular model had a high accuracy of prediction.

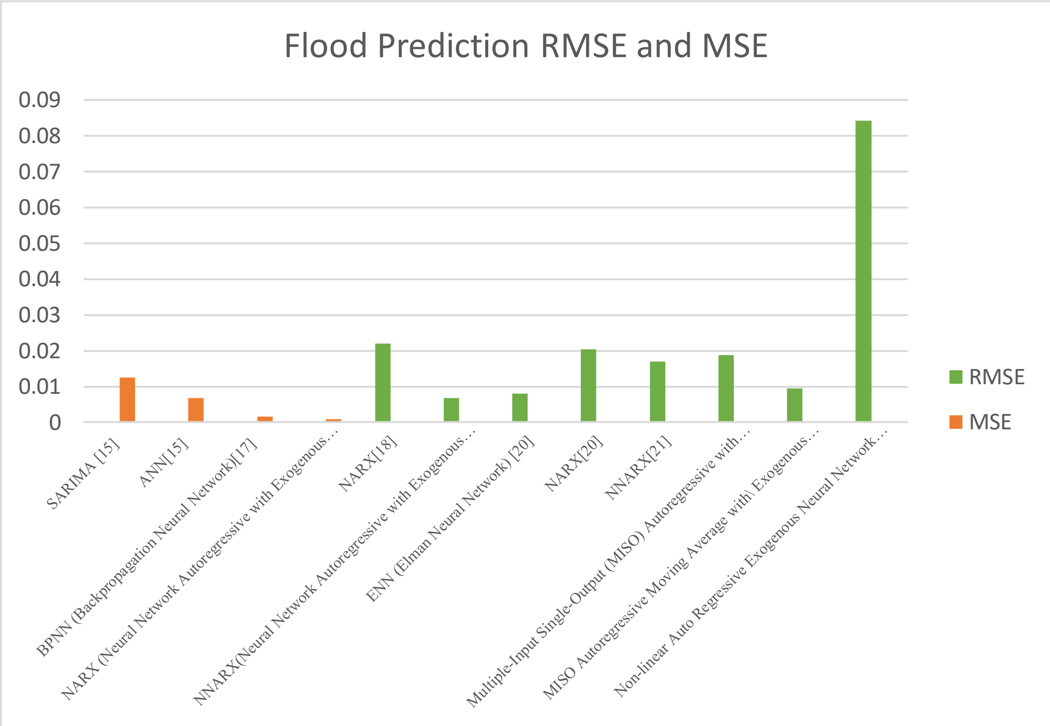

Figure 6 Flood Prediction RMSE and MSE

This situation difference compares to RMSE and MSE. The lower the value of RMSE and MSE mean the model is significant because smaller RMSE and MSE shows that the prediction graph is near to the mean, which indicates that the model is more accurate.

When deciding on the model to be used, it is crucial to know the purpose of the flood prediction model. There are two purposes in the flood prediction model, which is to predict the flash flood or seasonal flash flood. Multiple types of flood resource variables can use for flood forecasting, too, such as rainfall, streamflow, and water level.

There are different types of flood prediction models with different purpose and accuracy. It can observe that the most accurate model to be used is by using machine learning such as NARX. Different flood prediction models had different accuracy. Besides, it also depends on how long the prediction duration. The longer the prediction duration, the percentage of best fit also will decrease, and this will reduce the flood prediction model accuracy. This finding match with the finding from [22].

5. CONCLUSION

Based on the survey, the most accurate model to predict flash flood is by using Neural Network Autoregressive with Exogenous Input (NNARX), which can predict 7 hours of the incoming flood with the best fit of 95.29%. While for seasonal flood, the best flood prediction model to be used is also by NNARX, which has small MSE, as shown in Table 1. As the prediction duration increase, the best fit or accuracy of the prediction will decrease. This paper can be the guideline for researchers to find the most suitable flood prediction model. For future research, the scope of the survey may not be limited to Malaysia but also other flood-prone countries.

Acknowledgment*: The authors would like to thank and acknowledge the financial support from Kementerian Pengajian Tinggi (KPT) and the educational guidance from Universiti Teknikal Malaysia Melaka (UTeM).*

REFERENCES

1. M. I. Ibrahim, *"KPKT sediakan RM25.9 juta bendung banjir kilat*," *Berita Harian Online*, 2019.
2. *"Review of The National Water Resources Study (2000-2050) and Formulation of National Water Resources Policy-Volume 16-Melaka."*
3. R. R. A. N.S. Noor Haryantie, Z. Munirah, H. Nur Hasinah, *"Assessment Of Existing Evacuation Centres Using On Flood,"* 2016.
4. N. F. Zakinan, *"Banjir kilat di Melaka, 20 kenderaan tenggelam, 5 kawasan terjejas,"* Berita Harian, 01-Dec-2019.
5. R. A. Hamid, *"Melaka flash flood evacuees now at more than 1,000,*" New Staits Times, 07-Jul-2019.
6. E. Destro *et al.*, *"Coupled prediction of flash flood response and debris flow occurrence: Application on an alpine extreme flood event,"* J. Hydrol., vol. 558, pp. 225–237, 2018.
7. S. Alaghmand, R. bin Abdullah, I. Abustan, and B. Vosoogh, *"GIS-based river flood hazard mapping in urban area (a case study in Kayu Ara river basin, Malaysia),"* Int. J. Eng. Technol., vol. 2, no. 6, pp. 488–500, 2010.
8. S. Suhartono, *"Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model," J. Math. Stat.*, vol. 7, no. 1, pp. 20–27, 2011.
9. E. Cadenas, W. Rivera, R. Campos-Amezcua, and C. Heard, "*Wind speed prediction using a univariate ARIMA model and a multivariate NARX model,"* Energies, vol. 9, no. 2, pp. 1–15, 2016.
10. H. Liu, H. qi Tian, and Y. fei Li, *"Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction,"* Appl. Energy, vol. 98, pp. 415–424, 2012.
11. S. V. Kumar and L. Vanajakshi, *"Short-term traffic flow prediction using seasonal ARIMA model with limited input data,"* Eur. Transp. Res. Rev., vol. 7, no. 3, pp. 1–9, 2015.
12. T. Mai, B. Ghosh, and S. Wilson, *"Short-term traffic-flow forecasting with auto-regressive moving average models,"* Proc. Inst. Civ. Eng. Transp., vol. 167, no. 4, pp. 232–239, 2014.
13. Hyndman R.J & Athanasopoulos, *"Forecasting: principles and practice, 2nd edition,"* OTexts: Melbourne, Australia, 2018. [Online]. Available: https://otexts.com/fpp2/index.html. [Accessed: 18-Feb-2020].
14. S. H. Arbain, *"Time Series Methods for Water Level Forecasting of Dungun River in Terengganu Malaysia,"* Int. J. Eng. Sci. Technol., vol. 4, no. 4, pp. 1803–1811, 2012.
15. N. H. Ab Razak, A. Z. Aris, M. F. Ramli, L. J. Looi, and H. Juahir, *"Temporal flood incidence forecasting for Segamat River (Malaysia) using auto-regressive integrated moving average (ARIMA) modelling,*" J. Flood Risk Manag., no. 11, pp. 56–63, 2016.
16. S. H. Arbain and A. Wibowo, *"Neural networks based nonlinear time series regression for water level forecasting of dungun river,"* J. Comput. Sci., vol. 8, no. 9, pp. 1506–1513, 2012.
17. Q. A. Lukman, F. A. Ruslan, and R. Adnan, *"5 Hours ahead of time flood water level prediction modelling using NNARX technique: Case study terengganu,"* 2016 7th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2016 - Proceeding, no. August, pp. 104–108, 2017.
18. R. Adnan, A. M. Samad, Z. M. Zain, and F. A. Ruslan, *"7 hours flood prediction modeling using NNARX structure: Case study Kedah,"* Proc. - 4th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE 2014, no. March, pp. 433–437, 2014.
19. S. M. Zainorzuli, S. Afzal Che Abdullah, R. Adnan, and F. A. Ruslan, *"Comparative study of elman neural network (ENN) and neural network autoregressive with exogenous input (NARX) for flood forecasting,"* ISCAIE 2019 - 2019 IEEE Symp. Comput. Appl. Ind. Electron., pp. 11–15, 2019.
20. F. A. Ruslan, A. M. Samad, and R. Adnan, *"Modelling of flood prediction system using hybrid NNARX and Extended Kalman Filter,"* Proc. - 2017 IEEE 13th Int. Colloq. Signal Process. its Appl. CSPA 2017, no. March, pp. 149–152, 2017.
21. F. A. Ruslan, K. Haron, A. M. Samad, and R. Adnan, *"Multiple Input Single Output (MISO) ARX and ARMAX model of flood prediction system: Case study Pahang,"* Proc. - 2017 IEEE 13th Int. Colloq. Signal Process. its Appl. CSPA 2017, no. March, pp. 179–184, 2017.
22. A. Faruq, S. S. Abdullah, A. Marto, M. A. A. Bakar, S. F. M. Hussein, and C. M. C. Razali, *"The use of radial basis function and non-linear autoregressive exogenous neural networks to forecast multi-step ahead of time flood water level,"* Int. J. Adv. Intell. Informatics, vol. 5, no. 1, pp. 1–10, 2019.