Review of: "Flood Prediction Using Artificial Neural Networks: A Case Study in Temerloh, Pahang"

Nithila Devi Nallasamy[1]

1 Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ

Potential competing interests: No potential competing interests to declare.

I would like to appreciate the authors for investing their time in an important challenge: monitoring the floods in the Global South. It is very much needed in countries that are more susceptible and vulnerable to flood damages. It is interesting to employ machine learning methods to detect floods, and an accurate and timely early warning system can save lives and properties. The authors have used a classification approach to identify whether the weather conditions would lead to flooding on a daily basis.

In addition to considering the machine learning aspects, the**hydro-meteorological and hydrological aspects** also need to be considered for formulating a robust early warning system. I would encourage the authors to rigorously improve their modelling framework based on the concerns that I have listed below,

1) Flood detection in this paper is done whenever the rainfall data is available from the stations. Nonetheless, from a practical point of view, a more effective use of machine learning in this context would be to employ the algorithm with a **sufficient lead time such as 1 or 2 days before using rainfall forecasts (regional weather models such as WRF or downscaled or bias-corrected global forecasts)**. The rainfall and streamflow (discharge) records at the stations for the previous days need to be used (eg, t-1, t-2, ….). Considering the previous days' information is crucial so as to bridge the uncertainty in the forecast data. Prediction with sufficient lead time would provide enough time for formulating rescue and relief operations, and eventually, the flood detection framework will be more meaningful.

2) I assume that since the authors mention that they use water-level data, they define a flood when the water starts to overflow the banks. I couldn't find any description where they mention this. Again, from the application perspective, it is important to systematically see when the bank overflow happens, what sort of damage is caused, and there needs to be more clarity in this aspect. The authors could work on **the ways to define flooding** by considering the spatial characteristics of the flood within the entire catchment and also the damages caused.

3) It is well-known that the machine learning techniques like ANN that have been used in this paper are data driven. Therefore, spatially and temporally the **data that is used for training and validation of the model should be exhaustive** if one is looking at establishing the model for flood prediction.

-**spatial context**: one needs to know the catchment extent (that contributes to the accumulated runoff) and mainly consider the weather stations within the boundary (in addition to the stations outside the catchment). In this paper, there are no figures explaining this rationale or the catchment boundary.

-**temporal context**: there can be different flood-driving mechanisms that are specific to the region considered. The floods can be due to cyclones, depressions, etc. Therefore, it is a safe bet to consider a longer duration of data of a decade or more than tens of events for training the model. If the water-level data is available only for a few years, then the weather and discharge data can be used for flood identification.

4) Having discussed all these points, the authors need to work on the**novelty of the approach**. What is it that the authors are contributing: in the prediction framework (algorithms used) or in the application? There needs to be a more systematic understanding of the hydrology of the area; otherwise, the data-driven model is not of much use in the longer run.

Please re-read the literature considering these pointers and reframe the modelling framework, and if possible, the objective too. I have also listed below a few studies,

1. Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia - ScienceDirect
2. Data-driven real-time prediction of urban floods with spatial and temporal distribution - ScienceDirect
3. Adaptive selection and optimal combination scheme of candidate models for real-time integrated prediction of urban flood - ScienceDirect