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

Miguel A. Zuniga-Garcia

Potential competing interests: No potential competing interests to declare.

Writing observations:
For example, in the abstract, there is a part that says: "Despite this....", "Despite" is used wrongly; it fits better with "Because" or something similar.

In the Materials and Methods section, it says "To achieve all three objectives of this research" when no objectives were described before.

Content observations:
The Data Preprocessing section should be more elaborated; it only says that they integrated all datasets and that they performed data transformation. Artificial Neural Networks require variables to be normalized from 0 to 1, but the authors do not mention that. Also, it would be good to see an example of the database they used for training; it would be enough to show only 5 observations, for example.

The Model Development section only describes the parameters of an ANN but does not mention which values the authors used. Also, only the names of the error metrics are mentioned, but they do not mention how any of them work, or if it is better to have a high or low value from the metric.

In the Results and Discussion, variables are used to form a correlation matrix, but variables are not described in detail before (I suggest an explanation of every variable in the Data Preprocessing section).

Considering the problem the authors are addressing, and the high cost it has to have a flood, I found it very concerning that 5 of 9 flood instances were classified incorrectly. They are using a highly imbalanced database; I suggest applying a technique like bootstrapping to tackle this problem.

Also, the authors say that the error is very low because MSE, MAE, RMSE are 0.009, 0.009, and 0.096, respectively. This result is misleading considering that 210 of the instances of the database are No flood and only 9 are flood. The model is classifying correctly the more dominant class in the dataset, and they try to minimize the fact that the most costly error is when the model predicts that there would not be a flood and actually there is one.