C303

Flash Flood susceptibility Prediction Using Machine Learning approaches in Hyper Arid Basins

○Mohamed SABER・Tayeb BOULMAIZ・Emad MABROUK・Karim ABDRADO・Sameh KANTOUSH・Tetsuya SUMI・Hamouda BOUTAGHANE・Mawloud GUERMOUI・Daisuke NOHARA

Introduction

Flash floods are one of the most hazardous disasters worldwide, especially in arid regions. One of the main challenges is real-time monitoring of flash floods. Within the last few years, data-driven approaches have been widely employed in various studies for flash flood susceptibility (FFS) mapping, therefore, the main objective of this work is to use the machine learning approaches to predict the flood susceptibility in Hurghada city, and its upstream wadi catchments along the Red Sea, Egypt. Hurghada city is an important touristic city, located on the Red Sea coast, with dense population distribution. It has been frequently affected by flash floods. The trend of annual rainfall based on satellite datasets is increasing in the area from 1983 to 2019 (Saber et al. 2020, and Abdrabo et al. 2020).

The first step for FFS mapping is to create the flood inventory map for the flood occurrences based on the observational flooded areas and historical records of the previous floods. In this paper, 890 points were identified for both flood and non-flood points. The selection of points was based on the observational floods and historical flood records as well as the inundations maps developed by rainfall -runoff modelling ((Saber et al. 2020, and Abdrabo et al. 2020). Multicollinearity analysis was conducted to evaluate the linear dependency between all factors in order to find high correlated variables.

Data processing and methods

In order to estimate the flash flood susceptibility for the target area, the database for input conditioning factors and flood inventory maps were prepared as shown in (Fig. 1).

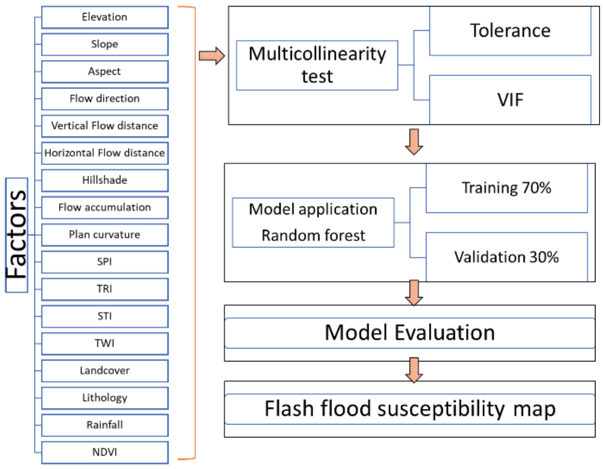

Fig. 1 Methodology flow chart for flash flood susceptibility.

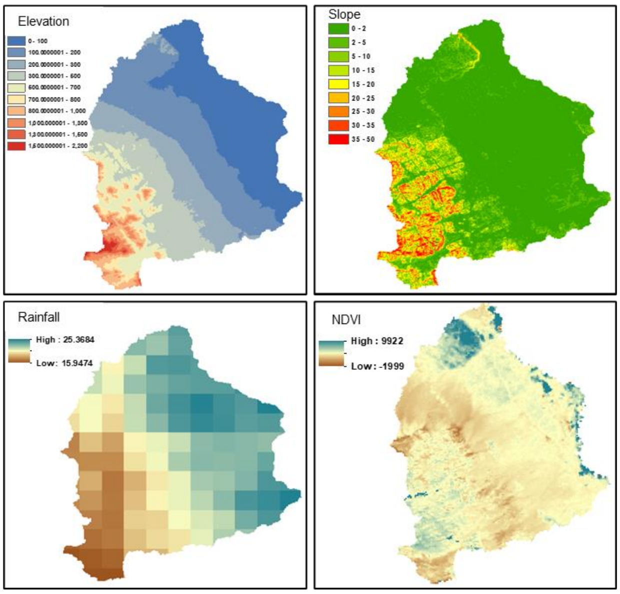

Fig. 2 Some of the selected flash floods conditioning factors

First the layers of variables were prepared using ArcGIS about 17 factors including (elevation, slope, Aspect, Plan curvature, vertical and horizontal flow distance, hillshade, flow direction, flow accumulation, topographic wetness index (TWI), stream power index (SPI), rainfall, lithology, land use/land cover (LU/LC), normalized difference vegetation index (NDVI), sediment transport index (STI), and topographic roughness index (TRI)). Examples of these layers are shown in (Fig. 2).

The input data in machine learning algorithms should be rescaled in the same range of values. Therefore, the statistical index coefficients were normalized between 0.1 and 0.9 by using the following equation:

$$y = 0.8 * \frac{x - \min(d)}{\max(d) - \min(d)} + 0.1,$$

where x is the original value of the variable, y is the standardized value of x, and d is the range value limits.

In this study, the machine learning approach of Random forest (RF) is used. It has been widely applied previously for FS mapping. RF is an ensemble learning method based on the decision tree model. The main steps in the implementation of RF methods are:

- Resample the data by generating equal sizes of subsets.
- Develop decision trees by using the subsets.
- Fuse the classification results of all developed decision trees by voting method.

After training of the RF model using 70% of the dataset, the remaining 30% was utilized to evaluate the model by using different performance tests such as Precision, Recall, F-score, and Area under Receiver Operating Characteristics (ROC) curve (AUC).

Results

After feature selection using correlation and multicollinearity analysis, RF model setting was tuned with a random search method applied on the number of trees and tree depth which intervals vary between [100 - 1000] and [2 - 20], respectively. The accuracy evaluation of RF model on the test dataset indicated satisfactory results translated by precision, recall and F-scores of 72.82%, 72.01% and 79.82%, respectively. The flash flood susceptibility map (Fig. 3) was developed by RF model using five classes: No, low, moderate, high and very high flood. The results of the FFS map shows that the high urbanized area of Hurghada city along the coast is located on both very high- and high-risk level, in addition to some areas sparsely distributed over the basin, especially on the upstream areas. The upstream desert basins are basically under the moderate to low level of hazard.

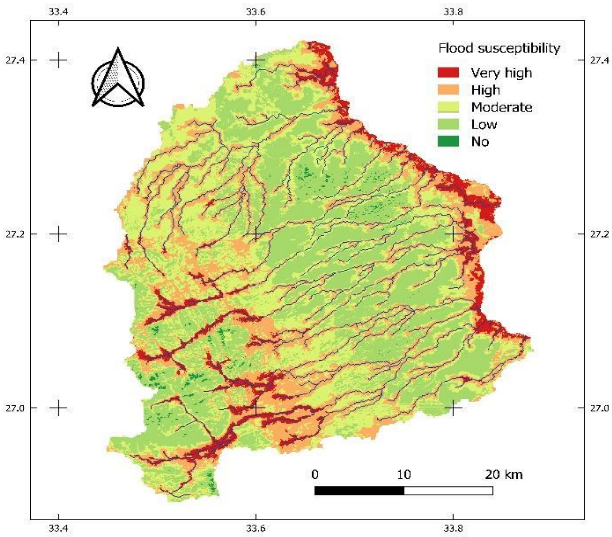

Fig. 3 Flash flood susceptibility map at the study area.

Conclusion

The results of this study can be significantly helpful for managers and decisions makers to mitigate the flash flood risks in such vulnerable- areas, especially in data-scarce regions.

References

Saber, M., Abdrabo, K. I., Habiba, O. M., Kantosh, S. A., & Sumi, T. (2020). Impacts of Triple Factors on Flash Flood Vulnerability in Egypt: Urban Growth, Extreme Climate, and Mismanagement. *Geosciences*, 10(1), 24.

Abdrabo, K. I., Kantoush, S. A., Saber, M., Sumi, T., Habiba, O. M., Elleithy, D., & Elboshy, B. (2020). Integrated Methodology for Urban Flood Risk Mapping at the Microscale in Ungauged Regions: A Case Study of Hurghada, Egypt. Remote Sensing, 12(21), 3548.