CONVOLUTIONAL NEURAL NETWORK-BASED PREDICTIONS OF POTENTIAL FLASH FLOOD HOTSPOTS IN SINGAPORE: INSIGHTS AND STRATEGIC INTERVENTIONS

JIONGYE LI[1] and RUDI STOUFFS[2]
[1,2]*Department of Architecture, National University of Singapore.*
[1]*lijiongye@u.nus.edu, 0009-0000-5496-0486*
[2]*stouffs@nus.edu.sg, 0000-0002-4200-5833*

Abstract. Amid increasing urbanization, changing climate, and limited stormwater infrastructure, urban flooding is a global issue, and Singapore is no exception. Traditional identification of flood-prone areas in Singapore has relied on historical flash flood data. However, by applying the booming influx of big data across various domains, including geography, weather, and DEM data, and using the deep learning model, Convolutional Neural Network (CNN), this research proposes a method that can accurately and effectively predict flash flood spots in an urban environment. Specifically, datasets including elevation, slope, aspect, rainfall, canals, drainage, and land use are fed into the CNN model to predict the locations of flash floods. The model, with a testing accuracy of 0.962, generates a comprehensive flash flood assessment map identifying high-risk areas in Singapore. Contrary to the current flood-prone area identification, which classifies only 0.79% of the country as susceptible to flash floods based on historical events, our CNN model-based assessment indicates that 11.4% of the country is at high risk. These newly identified zones are predominantly located along the coastline and in low-lying watershed outlets. Additionally, we propose corresponding stormwater infrastructure enhancements to mitigate flash flooding in these locations.

Keywords. Flash Floods, Flood Prediction, Convolutional Neural Network, Geospatial Data, Flash flood Assessment Map, Stormwater management measures

1. Introduction

In the context of escalating urbanization, shifting climatic conditions and inadequate stormwater infrastructural capacity, urban flooding presents a persistent global challenge, with Singapore being a prime example. This research delves into the factors contributing to urban flooding in Singapore and proposes a novel approach for predicting potential flash flood locations.

Firstly, monsoon rainfall patterns in Singapore frequently challenge the capacity of stormwater drainage infrastructure with intense, short-duration rainfall events (Chow

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 2, 69-78.

et al., 2016). Secondly, the inefficiency of drainage systems during heavy rainfall, particularly in low-lying areas, exacerbates flooding risks. These areas, characterized by high groundwater levels, struggle with stormwater infiltration, leading to water accumulation and subsequent flooding (PUB, 2022). Thirdly, rapid urbanization, driven by population growth and economic development, has transformed substantial green spaces into impervious urban landscapes. The decrease in forest cover from 35% to less than 10% in Singapore (Chow et al., 2016) has resulted in reduced ground infiltration and increased surface runoff, thereby heightening flood risks.

Historically, flood hotspots and prone areas in Singapore have been identified by the Public Utilities Board (PUB) based on past flooding events (PUB, 2021). However, these are reflective of historical rather than future flood risks. Advanced prediction of potential flash flood locations is crucial for guiding urban planners in implementing pre-emptive measures. Traditional flood mapping methods, such as linear logistic regression, have proven inadequate due to the complex nature of flooding (Jaafari et al., 2021). Although machine learning models like Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) have been employed in flood prediction (Choubin et al., 2019), their accuracy is limited due to the intricate interplay of natural input factors (Khosravi et al., 2019; Wang et al., 2019).

The advent of more sophisticated models, such as Convolutional Neural Networks (CNN), has been proposed due to their superior performance over traditional machine learning models (Sameen et al., 2020; Bai and Peng). The availability of big data across various domains enables CNN to process vast amounts of data, leading to highly accurate predictions (Ghorbanzadeh et al., 2019).

In addition, various stormwater management strategies, including Low Impact Development (LID), Water Sensitive Urban Design (WSUD), and Best Management Practices (BMPs), have been implemented globally for sustainable stormwater management (Fletcher et al., 2014). In Singapore, the Active, Beautiful, Clean Waters Programme (ABC projects) by the Public Utilities Board (PUB) incorporates infrastructures like bioretention swales and rain gardens to manage stormwater effectively (PUB, 2013).

This research aims to utilize advanced model algorithms and the availability of big data to accurately predict potential flash flood locations in Singapore. Based on the predicted flash flood map, it proposes relevant strategies to mitigate flood risks. The outcomes of this research could provide valuable insights for urban planners in optimizing stormwater management infrastructure deployment.

2. Study Area

This study selects Singapore as the research site due to its susceptibility to urban flooding. Singapore, a highly urbanized city-state, experiences a typical tropical climate (Meteorological Service Singapore). The tropical climate in Singapore is characterized by heavy and abundant rainfall throughout the year (Meteorological Service Singapore). According to rainfall records from 1981 to 2010 provided by the Meteorological Service Singapore, the average annual rainfall at the study site is 2165.9mm, with the highest monthly rainfall typically occurring in December, exceeding 200mm. Notably, due to the absence of detailed flood information and rainfall data for Tekong Island in the northeastern part of the country, this research excludes the island from its scope.

3. Methodology

3.1. RESEARCH METHOD DIAGRAM

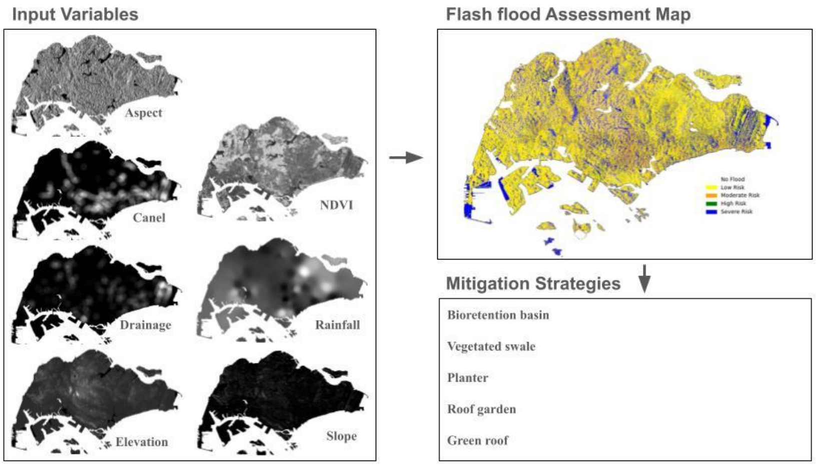

Figure 1. Overall research method diagram

3.2. INPUT AND TARGET VARIABLES

Per the literature reviews, seven key factors closely related to flash floods have been identified, including slope, elevation, aspect, locations of drainage, locations of canals, rainfall, and Landsat Normalized Difference Vegetation Index (NDVI) (Tehrany et al., 2014; Wang et al., 2019; Khosravi et al., 2019). The slope determines the rate of runoff and infiltration, while elevation plays a crucial role in deciding the species of plants covering the land. Aspect impacts soil moisture levels due to differences in solar radiation. The locations of drainage and canals are instrumental in determining the rate at which runoff is discharged. Rainfall is a critical factor as it decides the amount of runoff; notably, since flash floods are related to short but intense rainfall events, this research focuses on acquiring the highest rainfall data within a two-hour period each day in December, the month with the highest rainfall based on data from 1981 to 2010 (National Environment Agency, 2023). The NDVI index, which reflects land surface characteristics, affects the speed of runoff and the infiltration rate.

Regarding the data sources for these variables, slope, elevation, and aspect data are interpreted from Digital Elevation Model (DEM) data acquired from the National Aeronautics and Space Administration (NASA). QGIS is applied to process the DEM data into slope, elevation, and aspect data at a resolution of 30 meters. The locations of drainage and canals are compiled from the Public Utilities Board's (PUB) drainage reports and OpenStreetMap (PUB, 2023). The highest rainfall data over a two-hour period daily in December is sourced from the National Environment Agency (NEA), gathered from 65 weather stations. NDVI is calculated using bands 4 and 5 from the Landsat 8 dataset, which categorizes the country's land cover into three categories: sparse vegetation, medium vegetation, and dense vegetation. Below Table 1 summarizes input data sources and their specifications.

Inputs	Sources	Descriptions
Slope	NASA's DEM	Resolution: 30 m x 30 m
Elevation	NASA's DEM	Resolution: 30 m x 30 m
Aspect	NASA's DEM	Resolution: 30 m x 30 m
Drainage	OSM and PUB	Format: shapefile
Canal	OSM and PUB	Format: shapefile
Rainfall	NEA	Unit: highest rainfall in 2 hours
NDVI	Landsat 8	Resolution: 30 m

Table 1. Input variables and their data sources.

For the target variables, 91 flash flood spots and flood-prone areas have been identified by the Public Utilities Board (PUB) in their 2021 report. These identified areas, initially presented as polygons, are geoprocessed in this research into 957 point features for more precise analysis. These point features are then divided into two sets: 80% for training and 20% for testing purposes. To ensure compatibility with the Convolutional Neural Network (CNN) model used in the subsequent analysis, all input and target variables are converted into raster format, maintaining consistent dimensions across the dataset.

3.3. DATA PROCESSING

The slope raster is generated using the 'Slope' function provided by the Geospatial Data Abstraction Library (GDAL) within QGIS, utilizing DEM data sourced from NASA. Similarly, the aspect raster is calculated using the 'Aspect' method from the GDAL library in QGIS, based on the same DEM data.

Drainage and canal data are acquired and drafted in shapefile format. Their density is visualized using the 'Density Analysis' plugin in QGIS, which displays the density of canals and drainage within the study area in raster format. Rainfall data, obtained from the weather stations of the National Environment Agency (NEA), are denoted as points. These points are then converted into raster format using the 'Density Analysis' plugin in QGIS. The NDVI is calculated using the following formula (1):

$$\mathrm{NDVI} = \frac{NIR-R}{NIR+R} \tag{1}$$

In the calculation of the NDVI, the NIR represents the near-infrared band, and R is the red band from the Landsat 8 satellite data. The value range of NDVI spans from -1 to 1, with higher values indicating greater density of vegetation cover. Regarding the flood hotspots, the points indicating their locations are also converted into raster files using the 'Density Analysis' plugin in QGIS. This conversion facilitates a uniform format for analysis, ensuring consistency in data representation. All these datasets, including slope, aspect, drainage and canal density, rainfall distribution, and NDVI, are stored in a raster format, maintaining the same dimensions across all data types. This uniformity is crucial for the accurate processing and analysis of the data in the CNN model.

3.4. CNN-2D MODEL

In this research, the CNN-2D model is employed to predict the susceptibility of flash floods, due to its proven high performance in the geoscience field (Wang et al., 2019). Initially, seven rasters containing information on the input factors are converted into JPEG images. These single-channel images are then stacked to create a multi-channel image with seven channels, each representing a different factor: aspect, density of canals, density of drainage, elevation, NDVI, rainfall, and slope.

The multi-channel image, containing key flash flood factors, is divided into training (80%) and testing (20%) datasets for the CNN model. The target dataset is a flash flood assessment map created based on the density of existing flood hotspots. This map is also an image with five channels representing different levels of flood risk, categorized as follows: areas with 0-1 flash flood hotspots are identified as very low risk; 2-4 hotspots as low risk; 4-6 hotspots as medium risk; 6-11 hotspots as high risk; and areas with more than 11 hotspots as very high risk. The multi-channel target image is split into training and testing datasets in the same proportion as the input factor image.

The model's structure consists of six layers, excluding the input layer. The first layer is a Conv2D layer with 20 filters, each with a 3x3 kernel size, using the 'relu' activation function and 'same' padding. The second layer is a batch normalization layer. The third layer is a dropout layer with a rate of 0.5, followed by a fourth layer, a max pooling layer with a 2x2 pool size. The fifth layer is a Conv2D transpose layer with 15 filters, a 3x3 kernel size, and a 2x2 stride. The final layer is a convolutional layer with five filters using a softmax activation function. The model uses the Adam optimizer and categorical crossentropy as its loss function. It is trained over 20 epochs, aiming to minimize the loss value and maximize accuracy. The model's accuracy is evaluated using a testing target image and the specified formula (2):

$$\text{Accuracy} = \frac{\text{Correct prediction of pixels}}{\text{Total prediction of pixels}} \qquad (2)$$

where the formula measures the rate of accuracy by measuring the amount of accurate predicted pixels versus the total predicted pixels. The parameters settings for the model are detailed in the following Table 2.

Parameters	Settings
Convolutional kernel size	3x3
Number of convolution units	20
Max pooling kernel size	2x2
Number of epochs	20
Optimizer	Adam
Learning rate	0.001
Dropout rate	0.5

Table 2. Parameters settings of the model

3.5. ASSESSMENT MAP OF FLASH FLOOD

Once the CNN model is trained, this research inputs the multi-channel image, which comprises data on slope, elevation, aspect, locations of drainage, locations of canals, rainfall, and NDVI, into the model. Using the matplotlib library, a flash flood assessment map is then generated. This map differs from the initial flash flood hotspots map; it provides a comprehensive view of various levels of flash flood susceptibility across the entire study site. The metric used to determine the levels of predicted susceptibility to flash floods is consistent with the categorization applied in the flash flood hotspots raster used during the model's training phase.

4. Results

4.1. INPUT AND TARGET VARIABLES

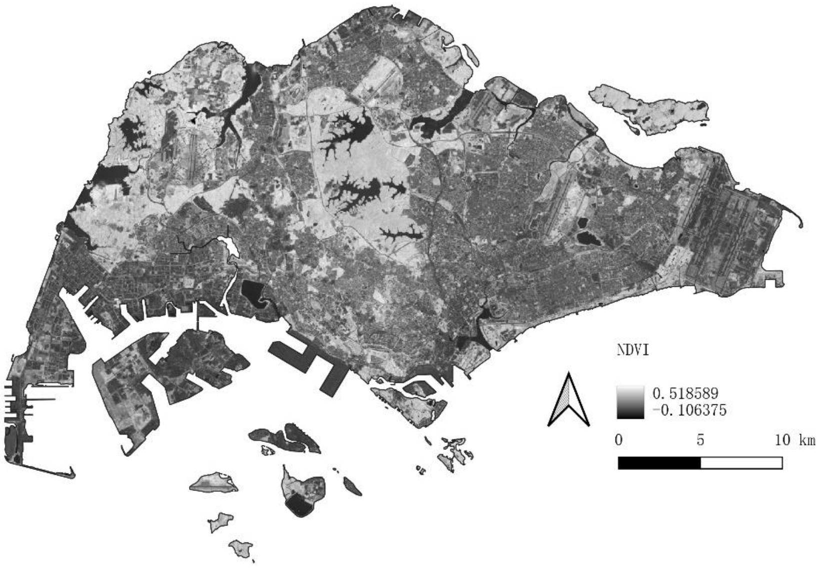

Figure 2. NDVI map

This study uses grayscale raster maps to represent variables, with lighter pixels indicating higher values. Figure 2 shows as an example of one of the input variables, the Normalized Difference Vegetation Index (NDVI). Figure 3 presents a heatmap of flash flood hotspots using hexagons in varying shades of grey. The hexagon tones range from darkest (8-31 hotspots) to white (no hotspots), providing a clear visualization of flash flood risk levels in the area.

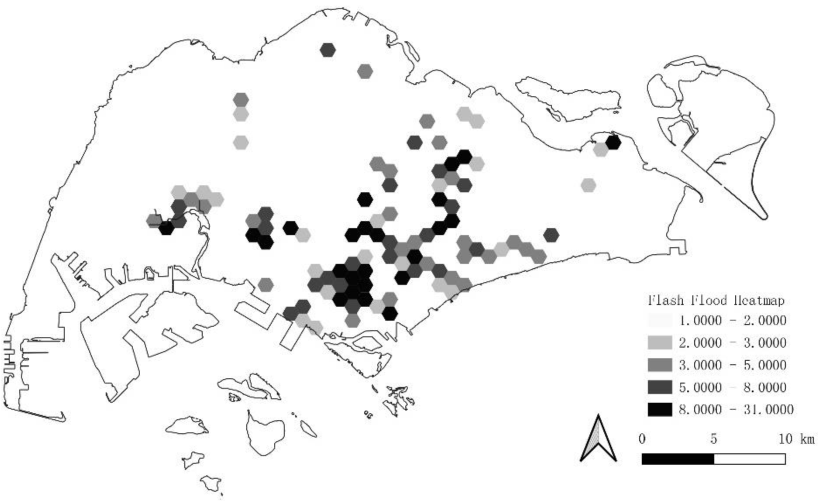

Figure 3. Flash flood hotspots map

4.2. FLASH FLOOD ASSESSMENT MAP

The trained model achieved an accuracy of 0.97, with a loss value of 1.28. The assessment map of flash floods predicted by the model is shown in Figure 4. The flood risk level on this map is related to the standards set for the target variable. In the target variable diagram, each hexagon measures 40 pixels by 50 pixels, totalling 900 pixels. The 'Severe Risk' level, represented by blue pixels, corresponds to the darkest hexagon in the target input diagrams. Each pixel in this category has a flash flood hotspot possibility ranging from 0.89% to 3.4%. The 'High Risk' level, depicted in green, indicates a possibility of flash flood hotspots for each pixel between 0.56% and 0.89%. The 'Moderate Risk' level, shown in orange, has a probability range for flash flood hotspots per pixel between 0.33% and 0.56%. The 'Low Risk' level, marked in yellow, corresponds to a flash flood hotspot possibility per pixel between 0.22% and 0.33%. The 'No Flood' level, represented by white colour, indicates areas where the possibility of flash flood hotspots per pixel is less than 0.22%.

Additionally, the map calculates the percentage of each risk category across the entire study site. Specifically, 11.4% of the total site's pixels are in the 'Severe Risk' category, 5.7% are in the 'High Risk' category, 41.6% are in the 'Moderate Risk' category, 35.1% are in the 'Low Risk' category, and 6.1% fall into the 'No Flood' category.

5. Discussions and Potential Strategies

The current method employed by PUB for flash flood estimation, focusing only on previously flooded low-lying areas, identifies a mere 0.79% of the country as flood-prone. In contrast, our CNN model-based flash flood assessment map reveals that 11.4% of the study area falls into the 'Severe Risk' category for flash floods, as depicted in Figure 4. These newly identified high-risk areas are mainly concentrated along coastal lines and in low-lying watershed outlets, highlighted by the red box in Figure 4. The map's prediction of 'Severe Risk' areas allows for the proposal of targeted measures to reduce the occurrences of flash floods. In this research, with a focus on sustainable infrastructure, the potential for implementing green infrastructure is explored, as opposed to relying solely on hard infrastructure such as canals, drains, barriers, etc.

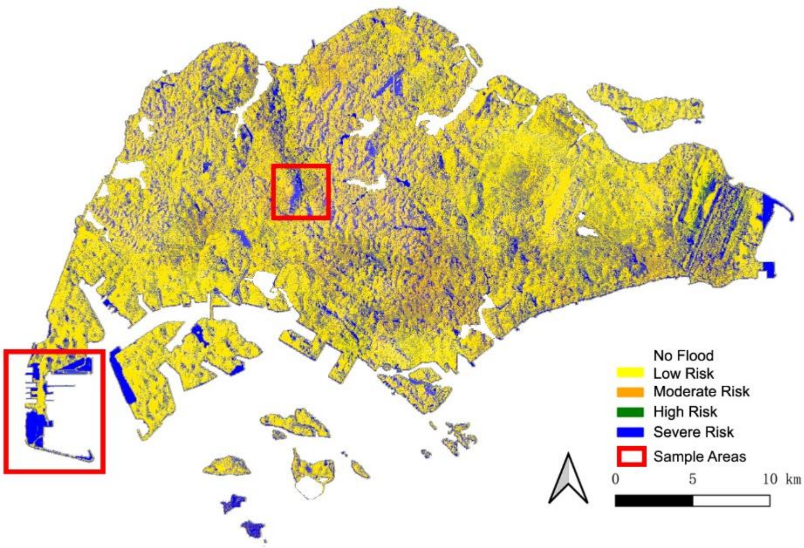

Figure 4. Flash flood assessment map

In the context of Singapore, a variety of green infrastructures, including rain gardens, bioswales, planters, and bioretentions, have been incorporated into the Active, Beautiful, Clean Waters (ABC) projects to effectively manage stormwater (PUB, 2013). However, these infrastructures have diverse requirements and are chosen based on the specific conditions of each site (PUB, 2013). Drawing from PUB's Drainage Handbook and previous research papers, this study summarizes the conditions for implementing potential stormwater mitigation measures in Singapore in Table 3 (PUB, 2013; Li et al., 2019; Fletcher et al., 2014). This summary provides insights into the suitability and effectiveness of various green infrastructure options in the context of the identified high-risk areas.

To mitigate potential flash flood risks identified by the assessment map, appropriate green infrastructure measures can be implemented at or near these locations. However, precise stormwater management also requires analysing additional factors such as building density, green space density, watershed, flow direction, and road locations. For instance, bioretention basins need ample ground-level space, making them suitable for areas with low building density and abundant green spaces. Their optimal placement is often near watershed outlets. Vegetated swales are typically aligned with roads, so road locations are crucial for their placement. In contrast, roof gardens, green roofs, and planters, which require less space, are suitable for areas with high building density.

Infrastructures	Conditions	Implemented Projects
Bioretention basin	Open space	Balam Estate
Vegetated swale	Open space; along road	Margaret Drive
Planter	Limited space	Khoo Teck Puat Hospital
Roof garden	Limited space; on buildings	Orchard Central Mall
Green roof	Limited space; on buildings	Orchard Central Mall

Table 3. Summary of stormwater management infrastructure

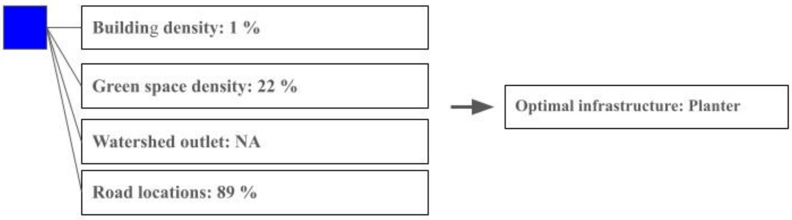

Figure 5. An example of how specific type of infrastructure can match a severe flash flood pixel

By overlaying all this information, including severe risk flood hotspots and the various aforementioned indices, the most appropriate stormwater infrastructure for each location can be determined. For example, a pixel in Figure 5 shows a low building index and close proximity to roads, suggesting the suitability of a vegetated swale. Future steps in this research will involve gathering detailed information on building density, green space density, watershed, flow direction, and road locations to further refine stormwater management strategies at high-risk flood locations.

6. Conclusions

This research addresses the severe issue of flash flooding in Singapore, intensified by rapid urbanization and land use changes. A novel method is developed, utilizing a range of geospatial data to train a Convolutional Neural Network (CNN) model for predicting potential flash flood locations. Key data inputs include slope, elevation, aspect, drainage and canal locations, rainfall, and NDVI. The trained CNN model successfully generates a detailed flash flood risk assessment map for Singapore, categorizing areas based on flood risk levels. The map reveals that 11.4% of the study area is at severe risk of flash flooding. The study also proposes categorization methods for these high-risk areas to guide the implementation of targeted stormwater management infrastructures, effectively combining predictive analytics with practical urban planning solutions for mitigating flash flood risks.

References

Bai, Z., & Peng, C. (2023). Convolutional Neural Network (CNN) Supported Urban Design to Reduce Particle Air Pollutant Concentrations.

HUMAN-CENTRIC - Proceedings of the 28th International Conference on Computer-Aided Architectural Design Research in Asia: CAADRIA 2023 (pp. 505-514). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Chang, C. L., Lo, S. L., & Huang, S. M. (2008). Optimal strategies for best management practice placement in a synthetic watershed. *Environmental Monitoring and Assessment*, 153, 359-364. https://doi.org/10.1007/s10661-008-0362-y

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. *Science of The Total Environment*, 651, 2087-2096. https://doi-org.libproxy1.nus.edu.sg/10.1016/j.scitotenv.2018.10.064

Chow, W. T. L., Cheong, B. D., & Ho, B. H. (2016). A Multimethod Approach towards Assessing Urban Flood Patterns and Its Associated Vulnerabilities in Singapore. *Advances in Meteorology*, 2016, 1-11 https://doi.org/10.1155/2016/7159132

Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J.-L., Mikkelsen, P. S., Rivard, G., Uhl, M., Dagenais, D., & Viklander, M. (2014). SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage. *Urban Water Journal*, 12, 525-542. https://doi.org/10.1080/1573062X.2014.916314

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S., Tiede, D., & Aryal, J. (2019). Evaluation of Different Machine Learning Methods and Deep-Learning Convolutional Neural Networks for Landslide Detection. *Remote Sensing*, 11(2), 196. https://doi.org/10.3390/rs11020196

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H. L., Hong, H., Chapi, K., & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, 573, 311-323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Li, C., Peng, C., Chiang, P.-C., Cai, Y., Wang, X., & Yang, Z. (2019). Mechanisms and applications of green infrastructure practices for stormwater control: A review. *Journal of Hydrology*, 568, 626-637. https://doi.org/10.1016/j.jhydrol.2018.10.074

Sameen, M. I., Pradhan, B., & Lee, S. (2020). Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *CATENA*, 186, 104249. https://doi.org/10.1016/j.jhydrol.2018.10.074

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332-343. https://doi.org/10.1016/j.jhydrol.2014.03.008

Wang, X., Kinsland, G., Poudel, D., & Fenech, A. (2019). Urban flood prediction under heavy precipitation. *Journal of Hydrology*, 577, 123984. https://doi.org/10.1016/j.jhydrol.2019.123984