

PAPER • OPEN ACCESS

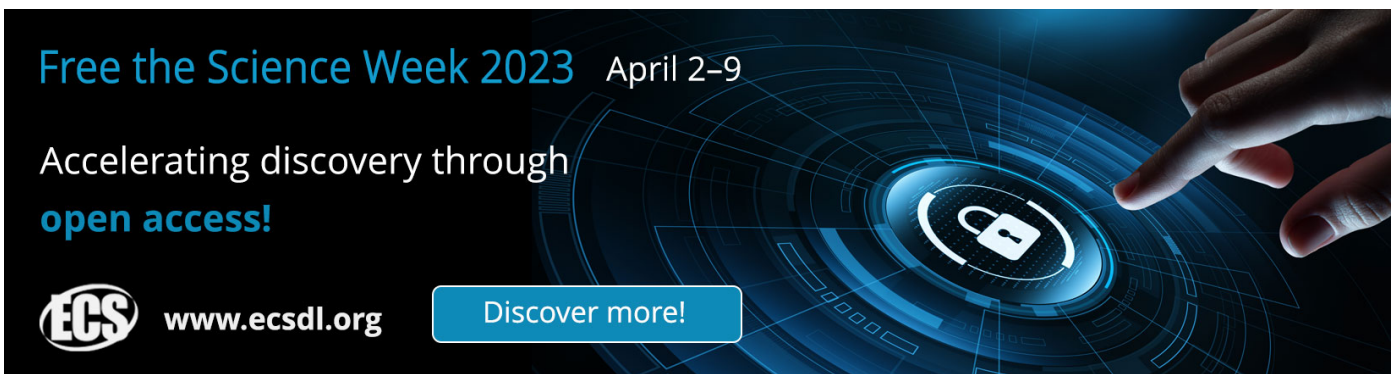
Flash Flood Susceptibility Mapping in Sungai Pinang catchment using Weight of Evidence

To cite this article: A Saleh *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1091** 012017

View the [article online](#) for updates and enhancements.


You may also like

- [Shape, size and dispersion of plant-driven silver nanoparticles for removal of methylene blue dyes](#)
N Isa, W Z Wan Kamis, V Inderan et al.
- [Pedestrian's perception toward quality of sidewalk facilities case study: UiTM Pulau Pinang](#)
Noorul Iqhlima Najwa Ismail, Nurul Aishah Abd Rahman, Noor Safwan Muhamad et al.
- [Feasibility study on potential of pedestrians' footstep based energy harvesting \(case study: UiTM Pulau Pinang\)](#)
Noorul Iqhlima Najwa Ismail, Nurul Aishah Abd Rahman, Ahmad Amzari Yaccob et al.



Free the Science Week 2023 April 2–9

Accelerating discovery through
open access!

 www.ecsdl.org [Discover more!](#)

The banner features a dark blue background with a futuristic, glowing blue interface. A hand is shown interacting with a circular element that contains a white padlock icon, symbolizing open access. The text is in white and light blue, with the ECS logo and website URL in white.

Flash Flood Susceptibility Mapping in Sungai Pinang catchment using Weight of Evidence

A Saleh¹, N Sabtu² and M R Bunmi²

¹ Disaster Preparedness and Prevention Centre (DPPC), Malaysia-Japan International Institute of Technology (MJIT), Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

² School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia

*Corresponding author e-mail: azlansaleh10@gmail.com

Abstract. Flash floods are one type of natural disaster that frequently occurs in Malaysia, especially in urban areas. This paper's objective is to develop the flash flood susceptibility mapping (FFSM) using weight of evidence (WOE) in the Sungai Pinang catchment. The Sungai Pinang catchment is located in the highly urbanised areas of Georgetown. The FFSM was developed using six flash flood conditioning factors: elevation, slope, rainfall, land cover, distance from the river, and lithology. The study showed that the WOE method is capable of producing high accuracy, with a success rate of 0.839 and a validation rate of 0.923 for the receiver operating characteristic's area under the curve (AUROC).

1. Introduction

Floods are a form of natural disaster that occurs on a regular basis in Malaysia. A flood can result in fatalities, environmental devastation, economic upheaval, and property loss. According to the World Meteorological Organization (WMO), flooding remains the third most devastating natural disaster, taking numerous lives and destroying destruction on infrastructure [1]. Direct and indirect destruction can both caused by flooding. Floods have the potential to directly result in the deaths of humans and animals, and property destruction. Indirect damage includes ecological changes, disease spread, and other implications due to floods [2].

According to Singh [3], floods affected roughly 178 million people, and the number of floods observed has grown dramatically during the last two decades. This is due to unmanaged development, a lack of drainage infrastructure and stream overflow, and extreme rainfall, which can result in flash floods in low-lying regions [4]. According to Dewan [5], the Balu River is used as an illustration of how development or land opening in low-lying areas has a substantial impact on future floods. This unplanned expansion has the potential to significantly increase runoff. Furthermore, because the main of Dhaka's urban districts are low-lying plain, elevation plays a crucial effect in the urban's flood threat. [5]. Dewan [5] added that altitudes between 1 and 4 metres are classified as high-hazard, while altitudes higher than 9 metres are classified as low-hazard. Individuals who live in low-lying locations or near rivers, for example, are in danger of floods. This demonstrates that elevation is a critical factor in determining an urban area's susceptibility to flooding.

Thus, the primary purpose of this work is to build the FFSM using the WOE model in accordance with six confounding variables: elevation, slope, rainfall, land cover, distance from the river, and lithology. To develop the FFSM, the current study will focus on the following objectives: (i) selecting flash flood conditioning factors (elevation, slope, rainfall, land cover, distance from the river, and lithology); (ii) analysing the correlation between the conditioning factors using a multi-collinearity test; (iii) developing the FFSM using WOE models; and (iv) determining model performance using the area under the receiver operating characteristic (AUROC) curve.



2. Study Area

The Sungai Pinang runs through Georgetown, the capital city of Penang. It is a developed and densely populated area [6]. The catchment of Sungai Pinang is just 3.1 kilometres long and spans an area of 50.97 kilometres square [7]. The study location is restricted by the longitudes of 100° 14' 26" East and 100° 20' 48" East, and the latitudes of 5° 26' 50" North and 5° 21' 29" North. Sungai Pinang was chosen for developing a flash flood susceptibility map (FFSM) because it is a low-lying area and susceptible to flash flooding. This study area is also the largest river catchment area on Penang Island with the highest population density. On 4 November 2017, a flash flood struck Penang. According to Davies [8], Kepala Batas meteorological station in Penang reported 946 mm of rain between 4 and 6 November 2017, with 458 mm of that amount falling in only 24 hours on 6 November 2017. The location of the study area is presented in Figure 1.

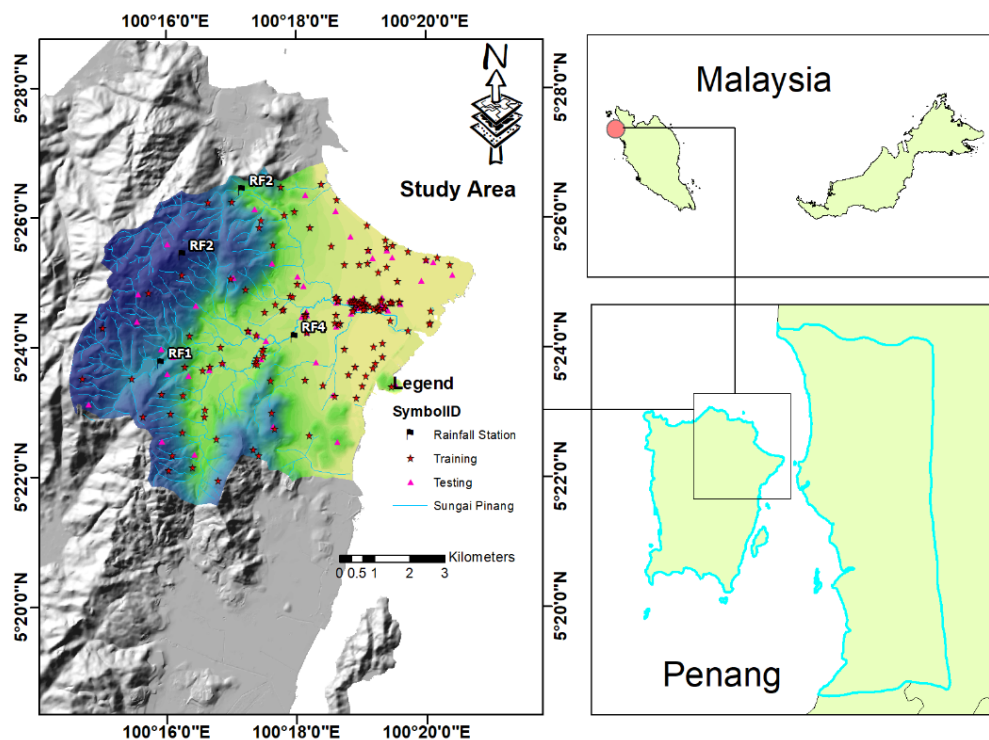


Figure 1. Location of study area Sungai Pinang catchment, Penang

3. Methodology

Firstly, a flash flood inventory will be created utilising data from the Penang GIS Centre (PEGIS) and the Majlis Bandaraya Pulau Pinang (MBPP) report. Based on the report, 110 flash flood points were recorded, which will be used in this study to develop FFSM. ArcGIS and Google Earth Pro were used to convert the report's inventory data to point data. Following data combination and validation, 70% of the entire data set was used for training and 30% for validation. This number is taken from past susceptibility mapping experiments, in which 70% of the time was spent on training and 30% on validation in the creation of FFSM [9–11].

Six conditioning factors were chosen for the analysis of FFSM. These conditioning factors are elevation, slope, rainfall, land cover, distance from the river, and lithology. *Elevation* is a very important factor in developing FFSM. Due to the movement of water from higher altitudes to lower altitudes areas, lower altitude or ground are prone to flooding. [12]. Flooding is more likely to occur in low-lying regions than in higher altitude places [13]. *Slope* is critical during a flood occurrence. When the angle of the slope increases, the infiltration time reduces, resulting in floods [14]. Furthermore, slope affects infiltration, drainage density, runoff, and erosion of the soil [15]. *Rainfall* is the essential component that contributes to the formation of a flood. Due to the common association

of flash floods with heavy and rapid rainstorms, rainfall is critical in flash flood prediction [16]. The conditioning factor maps for this study are depicted in Figure 2.

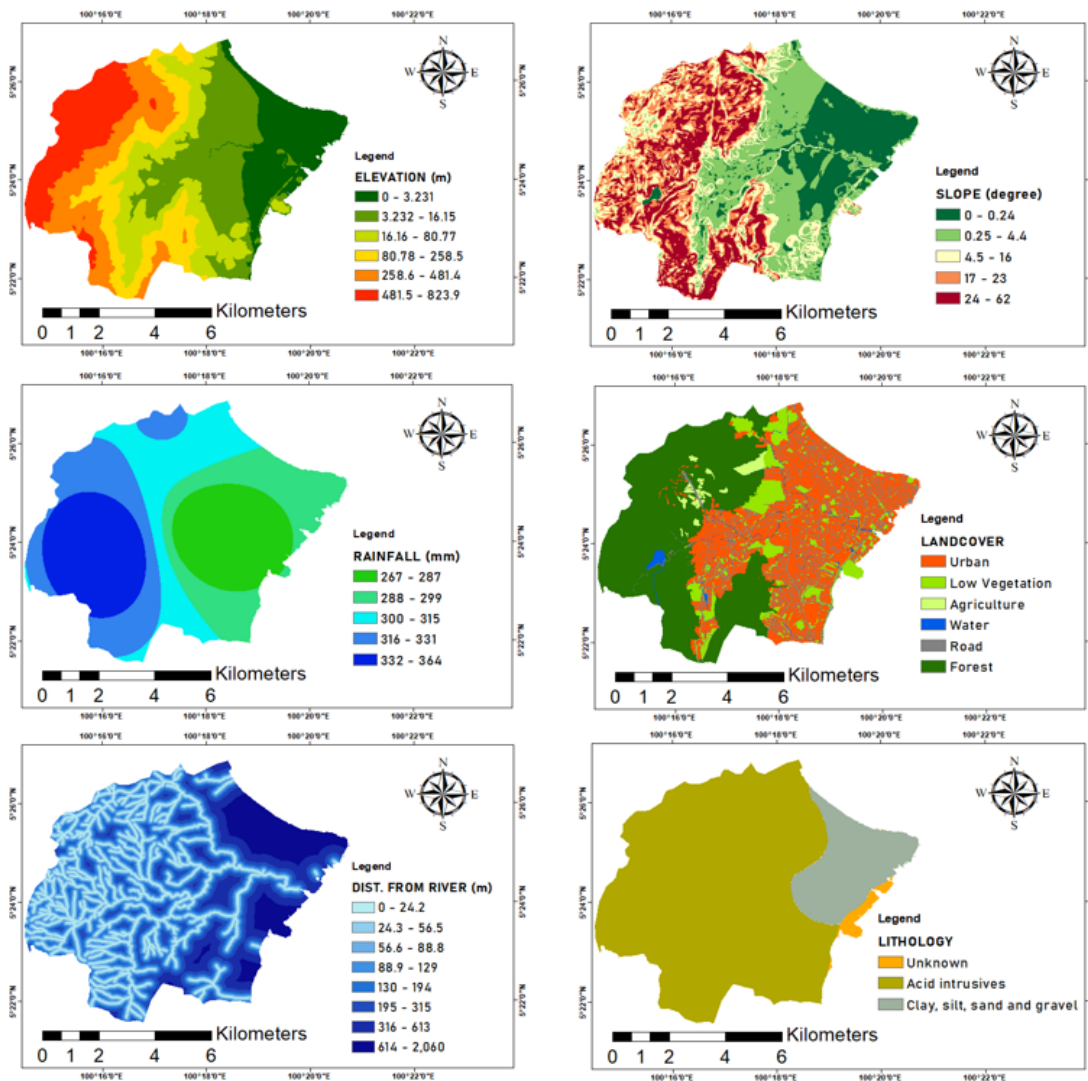


Figure 2. The conditioning factors of this study.

Land cover in FFSM is taken into account appropriately, since the dense vegetation and forest area coverage result in less surface runoff than developed regions. Land cover has a significant effect on runoff's velocity, interception, infiltration, and evapotranspiration [17,18]. *Distance from river* factor calculated in an ArcGIS environment using the Euclidean distance tool. It was revealed that decreasing the distance from the river component considerably increases the chances of flooding; increasing the distance significantly decreases the probability of flooding [19]. *Lithology*, the present study divided lithology into three groups by utilising a geological map obtained from PEGIS. (Unknown class, Acid Intrusive class, and Clay, Silt, and Gravel class).

Following the selection of the conditioning factors, multi-collinearity analysis was utilised to assess the correlation between them. This method was used by [10,20,21] to analyse the correlation between factors. Then, the FFSM will be developed after the analysis of the WOE method. The W_f 's values of conditioning factors will be assigned to each raster layer using the raster calculator in ArcGIS to develop FFSM. Finally, the FFSM will be analysed using the area under the curve of receiver operating characteristic (AUROC) to determine the accuracy of the FFSM. Figure 3 shows the flowchart of the methodology.

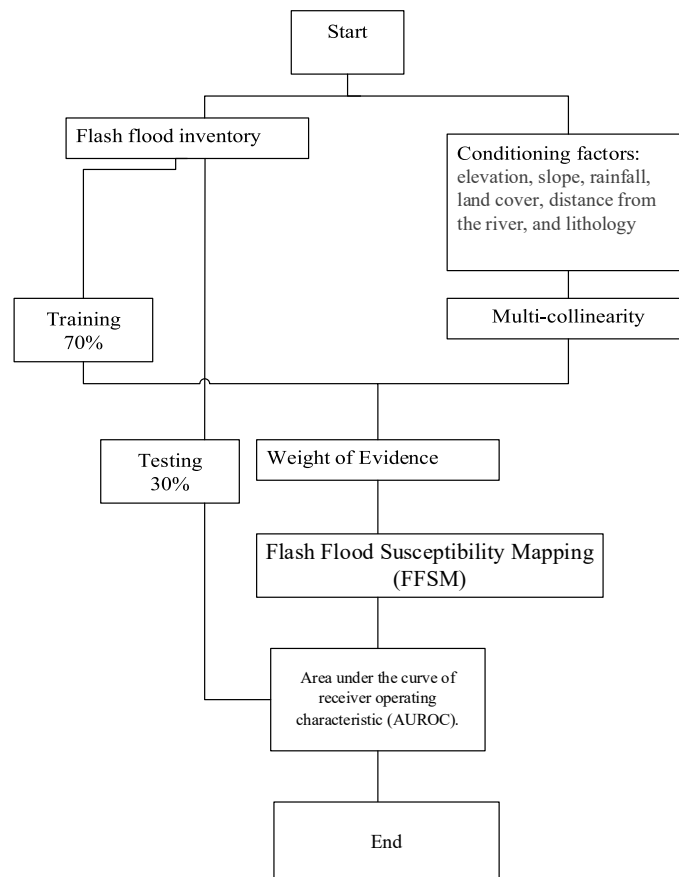


Figure 3. Flowchart of the methodology

3.1. Weight of Evidence (WOE)

In this study, the WOE method is used to compute the weights assigned to each variable category or class in relation to flash flood occurrences. As indicated previously by Costache et al. [22], in order to derive the WOE coefficients, the positive (W^+) and negative (W^-) weights must first be computed. Positive weights suggest a strong association between a variable category and flash flood spots or pixels, whereas negative weights suggest that this spatial relationship does not exist. Weights should be computed using equations (1) and (2) [23]:

$$W_i^+ = \ln \frac{P\{B|A\}}{P\{B|\bar{A}\}} \quad (1)$$

$$W_i^- = \ln \frac{P\{\bar{B}|A\}}{P\{\bar{B}|\bar{A}\}} \quad (2)$$

where P denotes the probability of a flash flood occurring and \ln represents the natural log. B and \bar{B} are denote the occurrence and non-occurrence of the variable, respectively. A and \bar{A} represent the occurrence and non-occurrence of flash floods [24]. The final WOE (W_f) value given to every variable category or class can be obtained as equation (3) [25]:

$$Wf = W_{plus} + W_{mintotal} - W_{min} \quad (3)$$

where W_{plus} indicates the positive weight of a class variable, W_{min} represents the negative weight of a class independent variable, and $W_{mintotal}$ denotes the total of all negative weights in an independent variable.

3.2. Multi-collinearity Analysis

When more than two predictor variables in a multiple regression model have a significant linear correlation, it is referred to as multi-collinearity in statistics [26]. In this study, multi-collinearity was identified using TOL and VIF. SPSS software was used to perform the multi-collinearity analysis. Khosravi et al. [27] previously reported that VIF greater than 10 or tolerance below 0.1 shows that there is a problem with multi-collinearity among the variables. Equations (4) and (5) were applied to determine VIF and TOL:

$$TOL = 1 - R_v^2 \quad (4)$$

$$VIF = \frac{1}{TOL} \quad (5)$$

where R_v^2 represents the coefficient of determination of an explanatory variable's regression on all other explanatory variables [28]. Multi-collinearity analysis is performed using SPSS software.

4. Results and Discussion

4.1. Multi-collinearity

Multi-collinearity is often not an issue or problem when the TOL (tolerance) value exceeds 0.1 and the VIF (variance inflation factor) number is less than 10 [29]. The TOL and VIF studies of this data set revealed no indication of multi-collinearity. The slope factor had the lowest TOL value of 0.322, while the distance from river had the greatest TOL value, which was 0.910. The highest VIF value observed in this current study is slope (VIF=3.101). The findings suggest that all of the conditioning factors can be used to generate the FFSM since their values fall within the specified range (TOL > 0.1 and VIF < 10). Table 1 presents the results of multi-collinearity analysis for this study.

Table 1. Multi-collinearity analysis

Conditioning factor	Collinearity Statistics	
	TOL	VIF
Elevation	0.452	2.213
Slope	0.322	3.101
Land cover	0.533	1.877
Rainfall	0.522	1.915
Distance from River	0.910	1.099
Lithology	0.723	1.384

4.2. Weight of Evidence

Table 2 shows the spatial relationship between flash flood occurrences and variables using the WOE method. The WOE method is used to determine the relationship between flash flood occurrences and classes of independent variables. Elevation factor values ranged from 1.40 in class 0-3.231 to 0.93 in class 3.231-16.155, 0.44 in class 481.431-823.925, 0.44 in class 80.777-258.486, 0.44 in class 258.486-

481.431, and -3.08 in class 16.155-80.777. It demonstrates that in the Sungai Pinang basin, the lowest elevation class has the largest effect on the occurrence of flash floods.

Based on Table 2, the greatest Wf value is 1.46 for the 0-1° slope class, followed by 0.96 for 0°, 14.20° (-0.66), 25-62° (-0.67), 1-5° (-0.98), and 5-14° (-1.58). The lowest Wf value is -2.96 for the 20–25° slope class. It demonstrates definitely that the low slope class is more vulnerable to flash floods. In terms of land cover, the road and urban classes have the highest weights of 0.97 and 0.88, respectively. Two components had a negative impact on flash floods: forest (-0.69) and agriculture (-0.14).

In terms of rainfall, the class 267–287 mm has the greatest Wf value (2.17), making it the sole class with a beneficial effect. The fourth class of this conditioning factor (316–330 mm) was discovered at a low Wf value of -3.84. The 88.9–129 m class had the greatest influence on the presence of flash floods, with a score of 1.028. There were just two classes where this conditioning factor was negative: 56.6-88.8 m and 88.9-129 m. The remaining classes had negative values, with the lowest being 130–194 m. The clay, silt, sand, and gravel classes all have a significant impact with 2.80 on flash floods in the Sungai Pinang basin. Acid Intrusive had a -1.85 value, whilst Unknown had a 0.24 value. Finally, a flash flood susceptibility map will be developed using ArcGIS software, using the WOE weights and equation 6:

$$FFSM_{WoE} = WoE_{elv} + WoE_{slope} + WoE_{landcover} + WoE_{rainfall} + WoE_{distriver} + WoE_{lithology} \quad (6)$$

Table 2. Weight of Evidence

Factor	Factor classes	W+	W-	Wf
Elevation	0 - 3.23	1.16	-0.51	1.40
	3.23 - 16.15	0.78	-0.42	0.93
	16.16 - 80.78	-2.62	0.18	-3.08
	80.78 - 258.49	0.00	0.16	-0.44
	258.49 - 481.43	0.00	0.16	-0.44
	481.43 - 823.93	0.00	0.16	-0.43
Slope	0	1.40	-0.07	0.96
	0 - 1	0.96	-1.02	1.46
	1 - 5	-0.40	0.06	-0.98
	5 - 14	-0.97	0.09	-1.58
	14 - 20	0.00	0.14	-0.66
	20 - 25	-2.32	0.13	-2.96
	25 - 62	0.00	0.15	-0.67
Land Cover	Urban	0.54	-0.46	0.88
	Low Vegetation	0.35	-0.05	0.27
	Agriculture	0.00	0.02	-0.14
	Water	0.14	0.00	0.01
	Road	0.91	-0.19	0.97
Rainfall	Forest	0.00	0.56	-0.69
	267 - 287	1.45	-1.56	2.17
	288 - 299	-0.46	0.09	-1.39
	300-315	0.00	0.21	-1.06
	316-330	-2.78	0.22	-3.84

	331-364	-2.02	0.19	-3.06
Distance from River	0–24.2	-0.12	0.02	-0.17
	24.3–56.5	-0.18	0.03	-0.24
	56.6–88.8	0.10	-0.02	0.09
	88.9–129	0.84	-0.22	1.03
	130–194	-0.19	0.02	-0.24
	195–315	-0.02	0.00	-0.06
Lithology	316–613	-0.10	0.01	-0.15
	614–2060	0.00	0.12	-0.15
	Unknown	-0.20	0.00	0.24
	Acid Intrusive	-0.93	1.36	-1.85
	Clay, Silt, Sand, and Gravel	1.43	-0.91	2.80

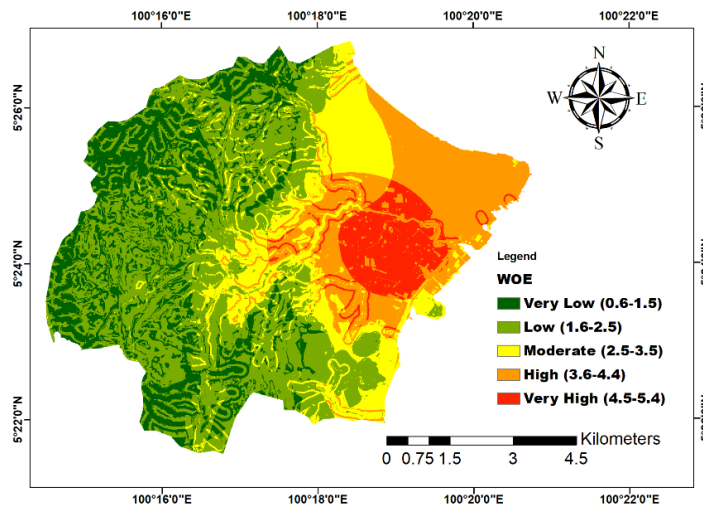


Figure 4. FFSM developed by WOE method

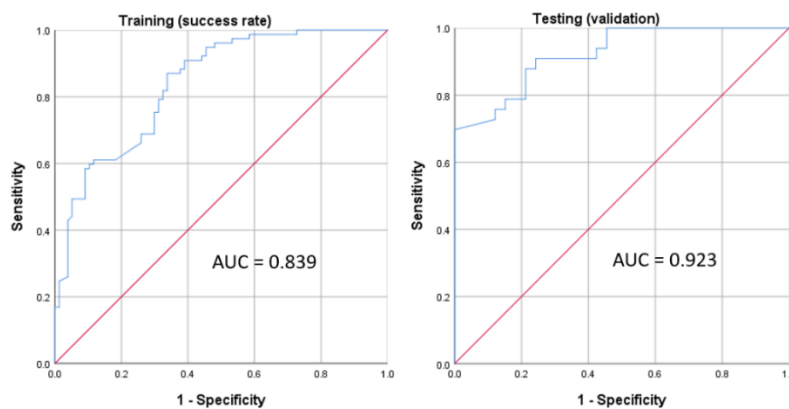


Figure 5. AUROC of WOE method

Following the development of the map using equation 6, it was divided into five susceptibility levels: very low, low, moderate, high, and very high. Figure 4 shows the FFSM using the WOE method. The success rate of the WOE method was 0.839, and the validation rate was 0.923. As Hong et al. [30] stated that the AUROC value can be divided into five classes, which are 0.5–0.6 (poor), 0.6–0.7 (average), 0.7–0.8 (good), 0.8–0.9 (very good), and 0.9–1 (excellent). Thus, an AUROC of 0.7 to 0.8 signifies a reasonable level of performance for the success rate. For the validation rate, it can be classified as excellent. The findings indicate that the WOE method can be used to create FFSM for this study area. Figure 5 shows the AUROC of the WOE method.

5. Conclusion

The WOE method was used in this work to create the FFSM for the Sungai Pinang catchment. The study considered six conditioning factors: elevation, slope, land cover, rainfall, distance from the river, and lithology. These components could be classified as standard or essential factors. The correlation between the conditioning factors was analysed using multi-collinearity analysis. These six conditioning factors demonstrate that there is no correlation and can be used to create an FFSM. Based on the FFSM, it indicates that the location near the major river of Sungai Pinang is the most vulnerable to flash floods. The AUROC of FFSM can be classified as good if it is greater than 0.7. The AUROC results show that the FFSM produced by the WOE model demonstrates very good performance for success rate and excellent validation rate. Thus, the FFSM of the WOE model can be used to help the local authorities manage and prevent flash floods in the study area. It is also important to give information about flash flood prone areas for future development and land use planning in the Sungai Pinang catchment areas.

Acknowledgments

The first author would like to thank his supervisor, Prof Ali Yuzir from Malaysia-Japan International Institute of Technology (MJIT) for his advices that improved the quality of the manuscript. The first author also would like to acknowledge MJIT and UTM for providing an Incentive Scheme to undertake his PhD program. The authors would also like to thank the MBPP, DID, JMG, PLANMalaysia, and PEGIS for providing the data for this study.

ORCID iDs

Azlan Saleh <https://orcid.org/0000-0002-7457-3114>

Nuridah Sabtu <http://orcid.org/0000-0003-0564-1272>

Mudashiru Rofiat Bunmi <http://orcid.org/0000-0003-3655-0074>

References

- [1] DID 2017 Flood Preparedness Guidelines
- [2] Koirala S 2011 Some flood damage estimate of the world *UNESCO-IHE Flood Manag. Educ. Platf.*
- [3] Singh D 2012 Unplanned urbanization increasing flood impacts *UNDRR*
- [4] Bhuiyan T R, Hasan Reza M I, Choy E A and Pereira J J 2018 Direct impact of flash floods in Kuala Lumpur City: Secondary data-based analysis *ASM Sci. J.* **11** 145–57
- [5] Dewan A 2013 *Floods in a Megacity: Geospatial Techniques in Assessing Hazards, Risk and Vulnerability*
- [6] Saad F N M, Rahman N N A, Kadir M O A and Omar F M 2006 Identification of pollution sources within the Sungai Pinang River Basin *Proceeding Malaysian Res. Gr. Int. Conf.* 478–85
- [7] Rahaman Z A, Farhana S, Rus C H E, Omar M A and Ismail W A N R 2016 Rivers and Lakes as Natural Heritage: Water Quality Status in the Northern States of Peninsular Malaysia **23** 109–28
- [8] Davies R 2017 Malaysia – Severe Storm and Floods Leave 7 Dead, 10,000 Displaced

FloodList

- [9] Saleh A, Yuzir A, Sabtu N, Abujayyab S K M, Mudashiru R B and Pham Q B 2022 Flash Flood Susceptibility Mapping in Urban Area using Genetic Algorithm and Ensemble Method *Geocarto Int.* **0** 1–27
- [10] Al-Juaidi A E M M, Nassar A M and Al-Juaidi O E M M 2018 Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors *Arab. J. Geosci.* **11** 765
- [11] Costache R, Bao Q, Avand M, Thi N, Linh T, Lee S, Nguyen D, Thi P, Nhi T, Vojtek M, Vojtekov J, Pham Q B, Avand M, Thuy Linh N T, Vojtek M, Vojteková J, Lee S, Khoi D N, Thao Nhi P T, Dung T D, Bao Q, Avand M, Thi N, Linh T, Lee S, Nguyen D, Thi P, Nhi T, Vojtek M and Vojtekov J 2020 Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment *J. Environ. Manage.* **265**
- [12] Khosravi K, Nohani E, Maroufinia E and Pourghasemi H R 2016 A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique *Nat. Hazards* **83** 947–87
- [13] Tehrany M S, Pradhan B and Jebur M N 2013 Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS *J. Hydrol.* **504** 69–79
- [14] Khosravi K, Pourghasemi H R, Chapi K and Bahri M 2016 Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models *Environ. Monit. Assess.* **188**
- [15] Conforti M, Aucelli P P C, Robustelli G and Scarciglia F 2011 Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy) *Nat. Hazards* **56** 881–898
- [16] Borga M, Anagnostou E N, Blo G and Creutin J 2011 Flash flood forecasting , warning and risk management : the HYDRATE project **4**
- [17] Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F and Mosavi A 2019 An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines *Sci. Total Environ.* **651** 2087–96
- [18] Yalcin A, Reis S, Aydinoglu A C and Yomralioglu T 2011 A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey *Catena* **85** 274–87
- [19] Talukdar S, Ghose B, Shahfahad, Salam R, Mahato S, Pham Q B, Linh N T T, Costache R and Avand M 2020 Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms *Stoch. Environ. Res. Risk Assess.* **34** 2277–300
- [20] Tehrany M S, Jones S and Shabani F 2019 Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques *Catena* **175** 174–92
- [21] Towfiqul Islam A R M, Talukdar S, Mahato S, Kundu S, Eibek K U, Pham Q B, Kuriqi A and Linh N T T 2021 Flood susceptibility modelling using advanced ensemble machine learning models *Geosci. Front.* **12** 101075
- [22] Costache R, Țîncu R, Elkhachy I, Pham Q B, Popa M C, Diaconu D C, Avand M, Costache I, Arabameri A and Bui D T 2020 New neural fuzzy-based machine learning ensemble for enhancing the prediction accuracy of flood susceptibility mapping *Hydrol. Sci. J.* **65** 2816–37
- [23] Van Westen C J, Rengers N and Soeters R 2003 Use of geomorphological information in indirect landslide susceptibility assessment *Nat. Hazards* **30** 399–419
- [24] Tehrany M S, Pradhan B and Jebur M N 2014 Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS *J. Hydrol.* **512** 332–43

- [25] Costache R, Blaschke T, Pham Q B, Pandey M, Arabameri A, Blaschke T, Pham Q B, Pham B T, Pandey M, Arora A, Linh N T T and Costache I 2021 Flash-flood potential mapping using deep learning, alternating decision trees and data provided by remote sensing sensors *Sensors (Switzerland)* **21** 1–21
- [26] Hong H, Liu J and Zhu A X 2020 Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble *Sci. Total Environ.* **718** 137231
- [27] Khosravi K, Pham B T, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I and Tien Bui D 2018 A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran *Sci. Total Environ.* **627** 744–55
- [28] Arora A, Pandey M, Siddiqui M A, Hong H and Mishra V N 2019 Spatial flood susceptibility prediction in Middle Ganga Plain: comparison of frequency ratio and Shannon's entropy models *Geocarto Int.* **0** 1–32
- [29] Saha S, Arabameri A, Saha A, Blaschke T, Ngo P T T, Nhu V H, Band S S, Thao P, Ngo T, Ha V and Band S S 2020 Prediction of landslide susceptibility in Rudraprayag, India using novel ensemble of conditional probability and boosted regression tree-based on cross-validation method *Sci. Total Environ.* 142928
- [30] Hong H, Duan H, Gigović L, Pamučar D, Wang Y, Chen W, Tien Bui D, Drobnyak S and Li S 2018 A Hybrid GIS Multi-Criteria Decision-Making Method for Flood Susceptibility Mapping at Shangyou, China *Remote Sens.* **11** 62