# Flash Flood Susceptibility Mapping of Sungai Pinang Catchment using Frequency Ratio
## (Pemetaan Kerentanan Banjir Kilat Tadahan Sungai Pinang menggunakan Nisbah Kekerapan)

AZLAN SALEH, ALI YUZIR* & NURIDAH SABTU

ABSTRACT

*Flash flood are natural disasters that frequently occur in Malaysia especially in urban areas. Due to this, the development of flash flood susceptibility mapping one of the tools used to aid the local authority in reducing and managing the flash flood impact. Frequency Ratio (FR) is a popular method in predictive modeling because of its capabilities to determine the critical conditioning factor of flash flood. The aim of this research was to compare the standalone FR with Ensemble FR-AHP. This ensemble method uses pair-wise comparison method between Frequency Ratio and Analytical Hierarchy Process (AHP). For this research, ten conditioning factors were selected which were slope, aspect, curvature, Topographic Wetness Index (TWI), Stream Power Index (SPI), Normalized Difference Vegetation Index (NDVI), distance from river, rainfall, elevation, and land use/land cover (LULC). The flash flood inventory was obtained from local authorities where the flash flood occurred in Penang, Malaysia on November 2017. 70% of 110 flooded locations were used as training dataset to assess the spatial distribution of flooding whereas the remaining 30% flooded locations were used as validation dataset. Based on the results, the prediction rate of FR-AHP method is slightly better accuracy compared to FR method which 88.33% (FR-AHP) and 85.62% (FR). The output of this research is crucial to assist local authority in land use planning and drainage system of the study area.*

*Keywords: Analytical hierarchy process; flash flood; frequency ratio; susceptibility mapping*

ABSTRAK

*Banjir kilat merupakan salah satu bencana alam yang kerap berlaku di Malaysia terutama di kawasan bandar. Disebabkan masalah ini, pembangunan peta rentatan banjir kilat boleh dijadikan alat untuk pihak berkuasa tempatan untuk mengurus dan mengurangkan risiko banjir. Nisbah Frekuensi (FR) adalah salah satu kaedah yang popular dalam ramalan model kerana kebolehupayaannya dan juga dapat menentukan faktor keadaan kritikal banjir kilat. Tujuan penyelidikan ini adalah untuk membandingkan FR dengan gabungan FR-AHP. Kaedah gabungan ini menggunakan kaedah perbandingan ikut pasangan antara Nisbah Frekuensi dengan Proses Hierarki Analitik (AHP). Untuk penyelidikan ini, sepuluh faktor keadaan dipilih iaitu cerun, aspek, kelengkungan, Indeks Kelembapan Topografi (TWI), Indeks Kuasa Aliran (SPI), Indeks Vegetasi Perbezaan Normalisasi (NDVI), jarak dari sungai, hujan, ketinggian, dan guna tanah/litupan tanah (LULC). Inventori banjir kilat diperoleh daripada pihak berkuasa tempatan - banjir kilat berlaku di Pulau Pinang, Malaysia pada bulan November 2017. 70% daripada 110 lokasi banjir digunakan sebagai data latihan untuk menilai taburan ruang banjir dan 30% lokasi banjir lain digunakan sebagai set data pengesahan. Berdasarkan hasilnya, kadar ramalan kaedah FR-AHP mendapat ketepatan yang lebih baik dibandingkan dengan kaedah FR iaitu 88.33% (FR-AHP) dan 85.62% (FR). Hasil daripada kajian ini dapat membantu pihak berkuasa tempatan dalam perancangan penggunaan tanah dan sistem perparitan kawasan kajian.*

*Kata kunci: Banjir kilat; nisbah frekuensi; peta rentatan; proses hierarki analitik*

## INTRODUCTION

Malaysia has recently experienced many devastating flash floods especially in urban areas due to increasing impervious area, poor drainage and planning; and high precipitation. According to Koirala (2011), to limit the damages resulting from floods, it is required to have the knowledge of factors resulting in flood forecasting, flood generation, estimation of flood discharges, flood hydraulics, possible effects of flooding, rainfall-runoff processes to help in acquiring the above information on flood and to limit the flood damages. In Malaysia, the government allocated approximately RM3.8 billion for the structural

flood mitigation projects for the entire Peninsular Malaysia, Sabah and Sarawak but this amount is only about 18% of the original request which needs to be implemented under the Ninth Malaysia Plan (Hussaini 2007).

In Malaysia, there are three types of flooding which are: flooding due to river bank overflow, high tides and flash floods (Zakaria et al. 2017). Heavy rainfall is the most common cause of flooding (Khosravi et al. 2018). Kia et al. (2011) stated that the 2006-2007 major flood in Johor was caused by abnormally heavy rainfall. However, the factor of flood occurs also depend on topographical, hydrological and geological (Tehrany et al. 2017). In Malaysia, the most recent major floods in 15 December 2014 - 3 January 2015 where more than 200,000 people were evacuated and moved to relief centres, while 21 were killed. This flood severely affected the East Coast and northern region of Peninsular Malaysia. The estimated total cost in terms of loss of property of these flood disasters amounted to $560 million and is considered as one of the costliest flood events in Malaysian history (Tjaardstra 2015).

The flood hazard mapping would produce the scale, frequency, location, and time of flood events. However, it is difficult to predict these things due to the limitation of the model and availability of data (Kia et al. 2011). Thus, the researchers have focused on flood susceptibility mapping instead of flood hazard mapping. Several methods have been utilized for flash flood susceptibility mapping (Saleh et al. 2020). This includes AHP (Youssef & Hegab 2019), Decision Trees (DT) algorithm (Khosravi et al. 2018), Shanon's Entropy (Khosravi et al. 2016a), Weight of Evidence (Costache 2019; Costache & Bui 2019; Khosravi et al. 2016b), Frequency Ratio (Cao et al. 2016; Youssef et al. 2016) and Logistic Regression (Youssef et al. 2016). Saleh et al. (2020) stated that the combination between GIS and analytical model can produce good accuracy in flash flood susceptibility mapping.

There is still lack of studies in flood susceptibility mapping in Malaysia, especially for flash flood. Researchers have developed flood susceptibility mapping in Kelantan (Pradhan & Youssef 2011; Tehrany et al. 2013), Terengganu (Tehrany et al. 2015) and Perlis (Dano et al. 2019). However, flash flood susceptibility mapping in Malaysia was not studied yet by any researcher. Flash flood susceptibility mapping not only produce the improved accuracy susceptibility mapping but it can also determine the critical or highest risk factor that triggers the flash flood. The aim of this research was to compare the capabilities of two methods which are the standalone FR and the Ensemble FR-AHP. This ensemble method uses pair-wise method consisting of Frequency Ratio and Analytical Hierarchy Process (AHP).

STUDY AREA

Sungai Pinang catchment was selected as the study area to develop the flash flood susceptibility mapping. Sungai Pinang catchment is located between longitudes 100° 15' 0" and 100° 25' 0" E and latitudes 5° 20' 0" and 5° 30' 0" N (Figure 2). This study area is located in north east of Penang Island, Malaysia. Penang population would be increase to about 1.8 million which is predicted by Department of Statistics Malaysia (The Star 2018). However, the number of available catchments in Penang is insufficient due to the lack of land. Flooding in Penang is typically caused by long and heavy rainfall due to insufficient of catchment to cater the heavy and long rainfall. This study area was selected for develop flash flood susceptibility mapping because Sungai Pinang was hit by several flash floods due to unpredictable, high intensity rainfall (climate change) combined with urban expansion and poor drainage systems. The urban expansion or land use changes in Sungai Pinang catchment area caused the increase of surface runoff which led to flash flooding in the area. Another reason is uncontrolled garbage dumping and wastewater discharges as mentioned by Chee et al. (2018), which causes blockage or clogged the drainage systems. The Sungai Pinang catchment is located at low-lying areas can also be one of the reasons this catchment is frequently exposed to flash flood vulnerability. The location of study area is presented in the Figure 1 which is located in Sungai Pinang, Penang, Malaysia. Thus, it is essential to develop a flash flood susceptibility framework to provide related decision supports to deal with such disasters.

METHODS

The methods of the study cover the steps taken to achieve the objectives and results of the study. The flow chart covering steps of this research's methods is shown in Figure 2.

Firstly, the flash flood inventory was acquired from the local authorities of Penang State Agencies which are Majlis Bandaraya Pulau Pinang (MBPP) and Sistem Maklumat Geografi Pulau Pinang (PEGIS). The flash flood inventory was validated by field visit using the Global Positioning System (GPS) and also using Google
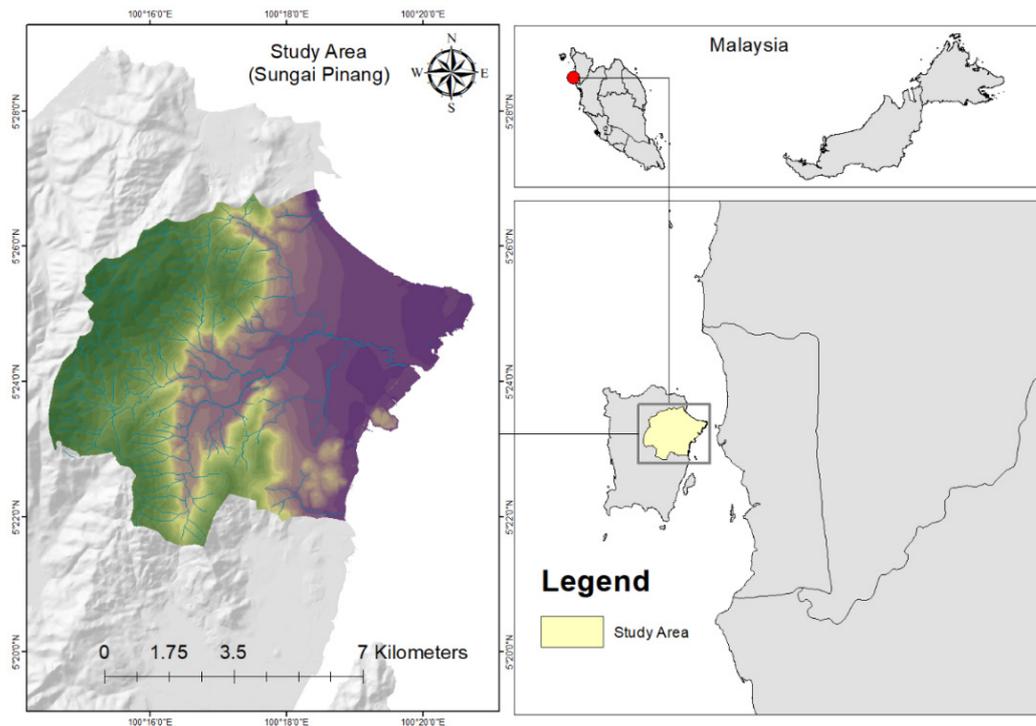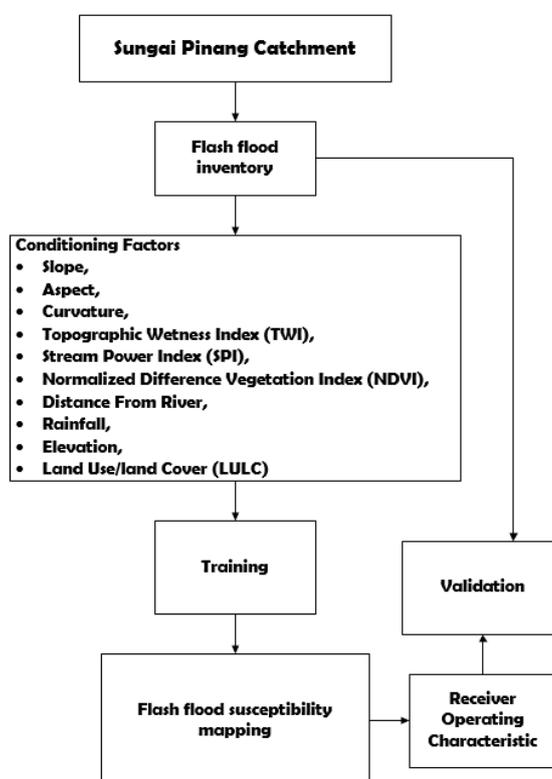
FIGURE 1. The location of study area



FIGURE 2. The flow chart of research methods

Earth Pro. Then, the flash flood inventory map was divided to 70% for training and 30% validation. This value is based on previous study of susceptibility mapping by previous researchers (Costache 2019; Rahmati et al. 2019; Samanta et al. 2018; Tehrany et al. 2014).

Then, the conditioning factor of flash flood would be determined first. The factors would be selected based on previous study and availability of data. Digital Elevation Model (DEM) would be used to extract altitude, slope, curvature, Stream Power Index (SPI), and Topographic Wetness Index (TWI). Landsat-8 data would be used to produce Normalized Difference Vegetation Index (NDVI) and Land Use Land Cover (LULC).

In the final step of the methods, the data of conditioning factors were sampled with flash flood points to develop the FR and FR-AHP model. The detailed

procedures would be discussed in Frequency Ratio method and Ensemble FR and AHP method. These two methods would be validated using Receiver Operating Characteristics (ROC) method.

DATA FOR FLASH FLOOD SUSCEPTIBILITY MAPPING

The basic set of information needed to develop a flash flood susceptible mapping. DEM is used to extracted slope, aspect, curvature, TWI, SPI, and elevation. Rainfall data was acquired from Department Irrigation and Drainage (DID) to generate rainfall map meanwhile data of distance from river was extracted from river shape file. For LULC and NDVI were produced by using Landsat-8 data. Table 1 shows the data for flash flood conditioning factors and source.

TABLE 1. Data for flash flood conditioning factors

| Type of data | Conditioning factor | Source |
|---|---|---|
| DEM | Slope, Aspect, Curvature, TWI, SPI, and Elevation | DID |
| Rainfall Data | Rainfall | DID |
| River | Distance from River | PEGIS |
| Landsat-8 | NDVI and LULC | EarthExplorer |
| Points | Flash flood points | MBPP & PEGIS |

FLASH FLOOD INVENTORY

According to Tehrany et al. (2014) and Youssef et al. (2016), for bivariate statistical analysis such as frequency ratio, the training data is set to 70% and for validation

model is 30%. The 70% of 110 flooded locations were used as training dataset to assess the spatial distribution of flooding and the other 30% flooded locations were used as validation dataset. Figure 3 shows the flash flood inventory map of the study area.
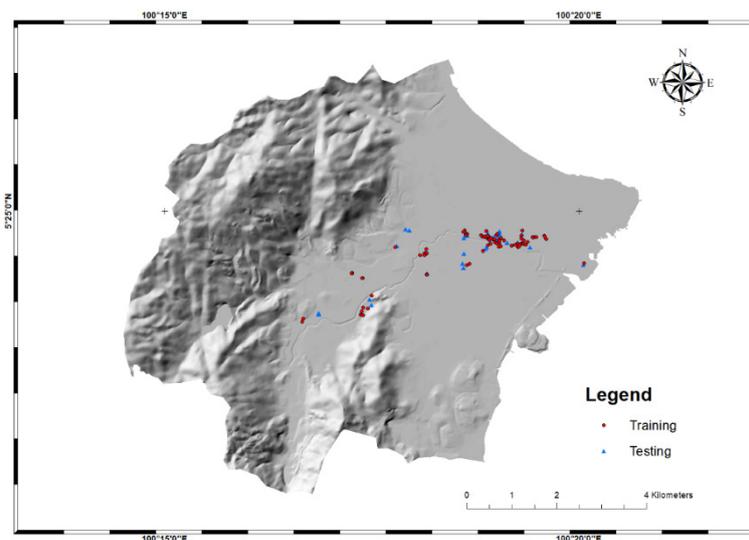


FIGURE 3. Flood inventory of study area

## FLASH FLOOD CONDITIONING FACTORS

According to Tehrany et al. (2019a), the flood conditioning factors is the most critical stage in constructing the final flash flood susceptibility maps and has the most significant influence on the accuracy of the output maps. Even though a lack of a structure on a selection of the flash flood conditioning factors, the most appropriate and frequently used flash flood conditioning factors by previous researchers (Bui et al. 2019a; Khosravi et al. 2018; Youssef & Hegab 2019) were used in the flash flood susceptibility model study. The thematic maps of flash flood conditioning factors of slope, aspect, curvature, TWI, SPI, NDVI, distance from river, rainfall, elevation, and LULC are presented in the Figure 4.

The slope is an essential factor in a flood event. As the slope angle increase, the time of infiltration decrease and runoff increases, resulting in floods (Khosravi et al. 2016a). Slope also affects the infiltration, drainage density, surface runoff, and soil erosion (Conforti et al. 2011). Elevation is also an important influencing factor to cause flash floods. Due to water flowing down from higher elevation, the low elevation regions have a high potential for flooding (Khosravi et al. 2016a). Distance from the river is also an essential factor in controlling the occurrence and flow of flood water as an area adjacent to the river is more prone to flooding (Khosravi et al. 2018).

Curvature factor was also derived from DEM. The curvature consists of three categories: concave, convex, and flat surface. It is a factor in runoff flow and can be useful in detecting flood susceptibility. Curvature effects flood events as topographic curvature affects runoff. Land use is considered in flood modelling because areas with high vegetation density and forest will reduce surface runoff in comparison to build up areas. The Landsat-8 data would be used to develop the current Land use map of study area. NDVI is a measure which describes the vegetation characteristic of an area, characteristics that influence both the surface runoff and infiltration capability of an area (Tehrany et al. 2013). Areas with less dense vegetation are considered more prone to floods (Tehrany et al. 2013). To generate NDVI, vegetation pixel data using Landsat-8 images are/were? developed before and after the flooding event. Normalized difference vegetation index (NDVI) is calculated using infrared and red bands of both datasets as shown in (1);

$$NDVI = \frac{(Band\ 5 - Band\ 4)}{(Band\ 5 + Band\ 4)} \qquad (1)$$

where Band 5 is reflectance unit of Near-Infra Red (NIR) band Landsat 8; and Band 4 is reflectance unit of red band Landsat-8 (Hashim et al. 2018). Reflectance is the ratio of the amount of light leaving a target to the amount of light striking the target. Rainfall is an important factor for this study because almost all flash flood occurred in Sungai Pinang catchment are due to heavy or high-intensity rainfall. For this study, the maximum of total rainfall on 4 November 2017 was selected to develop the rainfall map.

SPI is considered as an index of the erosive power of overland flow at the given point of the topographic surface; thus, it is one of the influencing factors for flood occurrences. It indicates that the power of flood is high when the SPI value is high and when the SPI value is lower, it shows the areas that have the potential of flow accumulation in the watershed (Chapi et al. 2017). As shown in (2), the SPI is calculated as:

$$SPI = A_s \tan \beta \qquad (2)$$

where $A_s$ is the specific basin area; and β is the local slope gradient (in degree). The high TWI values indicate areas susceptible to saturated land surfaces and areas that carry the potential to produce overland flow. The TWI factor has been used widely to characterize the effect of topography on the place and size of saturated source areas of runoff generation. As shown in the (3), the TWI is calculated as:

$$TWI = Ln \left( \frac{\alpha}{\tan \beta} \right) \qquad (3)$$

where α is the cumulative upslope draining area; and β is the slope gradient in degrees (Bui et al. 2019b). The highest values of TWI were mostly recorded in valley bottoms, terraced surfaces, and gentle slopes.

## FREQUENCY RATIO

Li et al. (2017) listed two reasons for the ongoing popularity of frequency ratio method. Firstly, the frequency ratio (FR) method is friendly to end users because of simplicity and accuracy. The ability to understand the input, calculation, and output procedures, as well as the ease of implementation on a GIS environment, makes the frequency ratio method be an acceptable simple tool of susceptibility assessment when sufficient data are available (Li et al. 2017; Pradhan et al. 2007). The second reason is the susceptibilities of a flood can be determined by the frequency ratio values
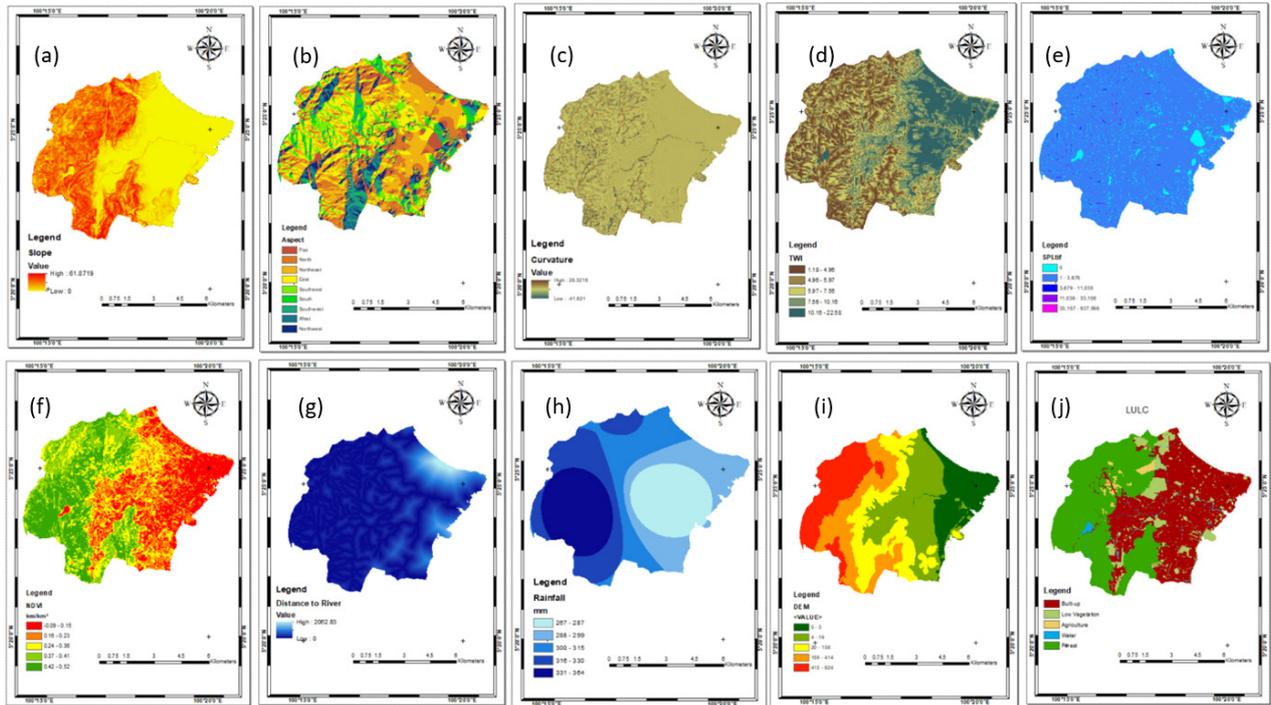
FIGURE 4. The thematic maps of flash flood conditioning factors of (a) slope, (b) aspect, (c) curvature, (d) TWI, (e) SPI, (f) NDVI, (g) distance from river, (h) rainfall, (i) elevation, and (j) LULC

calculated for each factor class. The frequency ratio method also can inspect the relationships between flood and flood conditioning factors. FR can be useful and can be applied to understand the quantitative relationship between flash flood points and flood conditioning factors, spatial datasets with different classes (Cao et al. 2016). This approach can be described as an FR index that represents the quantitative relationship between flash flooding occurrence or points and different conditioning parameters. It is stated based in (4):

$$FFHSI = \sum FR \qquad (4)$$

where FFHSI is the flash flood hazard susceptibility index and FR is the frequency ratio for each parameter (Jebur et al. 2015; Khosravi et al. 2016b; Tehrany et al. 2017). The FR can be defined as the ratio of the area where flash flooding hazards may occur to the total study area, or the ratio of the probability of a flash flood hazard occurrence to a non-occurrence as shown in (5):

$$FR = \frac{A/B}{M/N} \qquad (5)$$

where A is the number of pixels with a flash flooding

hazard for each class of each parameter; B is the total number of pixels with flash flooding hazards in study area; M is the number of pixels for each class of the parameter; and N is total number of pixels in the study area (Regmi et al. 2014; Tehrany et al. 2017; Youssef et al. 2016). In this analysis, if the FR value is greater than 1, it means there is a stronger correlation, whereas a value of less than 1 means there is a weaker correlation (Pradhan et al. 2011).

ENSEMBLE FREQUENCY RATIO (FR) AND ANALYTICAL HIERARCHY PROCESS (AHP)

For flash flood susceptibility mapping, the model needs many parameters to increase the accuracy of model. The objective of this ensemble is to assess the capability of FR-AHP model. This ensemble method is using pair-wise method between Frequency Ratio and Analytical Hierarchy Process (AHP) pair-wise comparison. This method was applied to increase the accuracy Frequency Ratio model. Khosravi et al. (2016a) proposed using the weights obtained from the FR model to apply as the input data for paired analysis instead of expert's opinions. The paired analysis which is also known as pair-wise method, was used before by Althuwaynee et al. (2014) and Ghosh et al. (2011) to develop landslide susceptibility mapping and later was used by Khosravi et al. (2016a) to

develop flash flood susceptibility mapping. The procedure of ensemble model of FR-AHP was adopted from Khosravi et al. (2016b). The procedure of this method is as follows: Firstly, equation 6 was used for rating each spatial factor with the training dataset.

$$PR = \frac{SA_{max} - SA_{min}}{(SA_{max} - SA_{min})_{Min}} \quad (6)$$

where SA is the index of the spatial association (FR) of spatial factors and flash flood, PR is the predictor rate. In Table 4, it shows that the value of the difference between the SA maximum and SA minimum was divided into minimum difference of total factors. The PR value shows the relationship between actual relative weights among the factors which is referring to actual the flash flood density.

For the next step, to determine the eigenvalue, individual numbers and the total number of the columns were divided. Integer values are more applicable to the rating, which can be obtained by dividing the fractional weight with the smallest weight of fractional weight. This fractional and integer weight of predictor are as shown in Table 4. The detailed procedure of this method can be referred to Althuwaynee et al. (2014) and Ghosh et al. (2011).

The final step of this method is to determine the CR value. The CR value is the ratio between consistency index of matrix and random index, the value of this CR ranges from 0 to 1. Althuwaynee et al. (2014) stated that the model would automatically be discarded when the CR value is larger than 0.1. It is stated based on (7):

$$CR = CI/RI \quad (7)$$

where RI is the average of the resultant consistency index and CI is consistency index, which is as follows in (8):

$$CI = (\lambda_{max} - n)/(n - 1) \quad (8)$$

where $\lambda_{max}$ is the largest eigenvalue matrix and n is the rank of matrix. FR values were normalized through dividing each number by its total amount of FR values to obtain the normalized value for ensemble model (Khosravi et al. 2016a).

## RESULTS AND DISCUSSION

### FREQUENCY RATIO

The relationship conditioning factor with floods were calculated as shown in Table 2. According to Tehrany et al. (2013), when the number of frequency ratio is high, it indicates the strong relationship between conditioning factors and flood occurrences. Khosravi et al. (2016a) and Pradhan et al. (2011) added that when FR value is more than 1, it indicates that there is strong relationship exist between flood and flood conditioning factor classes, and the relationship is weak when the FR value is less than 1. In the case of the slopes factor, the results showed that the slope class of 0° and 0°-1° were 4.76 and 2.68, respectively; and for slopes more than 1°, they were less than 1. The highest FR value (4.76) can be found in this conditioning factor which is corresponds to the 0° areas. Mostly flood occurred in 0° and 0°-1° class which is about 87% based on flood inventory. Based on the result, the lower slope class (0° to 1°) have more probabilities of flash flood and the higher slope would reduce the possibilities of flooding. For aspect factor, most classes of this factor shown a relationship with flash flood, where the strongest found at the flat class (2.28) and lowest is northeast class (0.28). For the flat class of aspect factor, the flash flood occurred is about 20% and for the lowest FR value of this factor is about 6% only.

For curvature, convex have the highest effect with FR value of 1.23, following the concave areas which 0.22 of the frequency ratio value. For this conditioning factor, FR value do not have a wide-ranging: from 0.13 for the flat class to 1.23 for convex. About 96% of the flash flood had occurred in the third class (convex class). In case of TWI, the first and second classes (1.18-4.96 and 4.96-5.97) have same value of FR which is 0.13. The highest value of FR for this conditioning factor is 10.16-22.58 class. Based on the result, it is shown that the higher value of TWI, would give significant impact on flash flood occurrences in the study area. Mostly flood occurred in these two classes (7.56 - 10.16 and 10.16 - 22.58) which are about 89%.

In the case of SPI, the highest value of FR is class 11035.27 - 33105.81 with 8.74 of FR value. For class 33105.81 - 937998, the value of FR is 0, it shows that this class did not give any impact on flood. For SPI, the second highest value is the first class (0) with 2.62 of FR value. For NDVI factor, the index values range from -0.093 to +0.516 which this factor was divided into 6 classes. The highest FR value is 2.63 which is recorded in 0.134-0.196 class. This result shows that the 0.134-0.196 class is high probability to flooding. Two classes obtained 0 value of FR which are 0.380-0.416 class and 0.416-0.516 class.

For distance from the river result, it's shown that the class of 82.46-120 m is the highest impact with FR value 2.09 followed by class of 284.12-557.52 m with FR value

of 1.16. Meanwhile, the class of 557.52 - 2062.83 m did not have any impact on the occurrence of floods. In the case of rainfall factor, the class of 267-287 mm which is the lowest rainfall factor has the highest frequency ratio (4.16). The first class (267-287 mm) of this conditioning factor corresponds to the maximum FR value of 4.16. Flooding happens mainly at lower elevations, according to Khosravi et al. (2016a), and heavy rainfall would not have a massive influence on flash flooding. Elevation has been another significant factor impacting floods. The lowest elevation class have the highest frequency ratio value which is 3.59 following the second lowest class of elevation (3.231 to 16.155). The first class of this factor (0 - 3.231) is the most susceptible areas to flash flood class which is about 56%.

One of the main conditioning factors for flash flood events is LULC. The growth of impermeable areas in urban areas will frequently cause flash floods. In the case of LULC, the maximum FR values are located in built-up areas of 1.96, followed by low areas of vegetation (1.29). However, the others classes which are agriculture, water bodies and forest areas have 0 value of FR. Based on the result, it is show that built-up areas and low vegetation areas were most susceptible areas to flooding compare the other LULC classes due to low impervious areas in that two classes. For this study area, the areas with lowest elevation, 0-1° or lowest slope angle, convex area, and urban areas are the most susceptible areas to flooding based on the FR analysis. Raster calculator tool in ArcGIS was used to develop the flash flood susceptibility mapping of FR model according to (9):

$$FSM_{FR} = FR_{slope} + FR_{asp} + FR_{curv} + FR_{TWI} + FR_{SPI}$$

$$+ FR_{NDVI} + FR_{river} + FR_{rain} + FR_{elev} + FR_{LULC} \quad (9)$$

TABLE 2. FR and PR estimated value for conditioning factor of study area

| Factor | Factor classes | No of points | % flood of points | FR | PR |
|---|---|---|---|---|---|
| Slope | 0° | 225 | 11.69 | 4.76 | 3.03 |
| | 0° - 1° | 1450 | 75.32 | 2.68 | |
| | 1° - 5° | 175 | 9.09 | 0.59 | |
| | 5° - 14° | 50 | 2.60 | 0.19 | |
| | 14° - 20° | 25 | 1.30 | 0.10 | |
| | 20° - 25° | 0 | 0.00 | 0.00 | |
| | 25° - 62° | 0 | 0.00 | 0.00 | |
| Aspect | Flat | 400 | 20.78 | 2.28 | 1.00 |
| | North | 200 | 10.39 | 0.64 | |
| | Northeast | 125 | 6.49 | 0.28 | |
| | East | 400 | 20.78 | 1.21 | |
| | Southeast | 225 | 11.69 | 1.11 | |
| | South | 175 | 9.09 | 1.52 | |
| | Southwest | 150 | 7.79 | 1.48 | |
| | West | 150 | 7.79 | 1.24 | |
| | Northwest | 100 | 5.19 | 0.83 | |
| Curvature | Concave | 50 | 2.60 | 0.22 | 3.68 |
| | Flat | 25 | 1.30 | 0.13 | |
| | Convex | 1850 | 96.10 | 1.23 | |

| | | | | | |
|---|---|---|---|---|---|
| TWI | 1.18 - 4.96 | 50 | 2.60 | 0.13 | 2.75 |
| | 4.96 - 5.97 | 50 | 2.60 | 0.13 | |
| | 5.97 - 7.56 | 100 | 5.19 | 0.24 | |
| | 7.56 - 10.16 | 675 | 35.06 | 1.80 | |
| | 10.16 - 22.58 | 1050 | 54.55 | 2.77 | |
| | | | | | |
| SPI | 0 | 350 | 18.18 | 2.62 | 3.78 |
| | 0 - 3678.42 | 1550 | 80.52 | 0.87 | |
| | 3678.42 - 11035.27 | 0 | 0.00 | 0.00 | |
| | 11035.27 - 33105.81 | 25 | 1.30 | 8.74 | |
| | 33105.81 - 937998 | 0 | 0.00 | 0.00 | |
| | | | | | |
| NDVI | -0.093 | 350 | 18.18 | 1.12 | 2.36 |
| | 0.134 - 0.196 | 875 | 45.45 | 2.63 | |
| | 0.196 - 0.308 | 675 | 35.06 | 2.07 | |
| | 0.308 - 0.380 | 25 | 1.30 | 0.08 | |
| | 0.380 - 0.416 | 0 | 0.00 | 0.00 | |
| | 0.416 - 0.516 | 0 | 0.00 | 0.00 | |
| | | | | | |
| Distance from River | 0 - 25 | 200 | 10.39 | 0.79 | 1.38 |
| | 25 - 53.15 | 275 | 14.29 | 1.13 | |
| | 53.15 - 82.46 | 175 | 9.09 | 0.73 | |
| | 82.46 - 120 | 500 | 25.97 | 2.09 | |
| | 120 - 176.14 | 275 | 14.29 | 1.15 | |
| | 176.14 - 284.12 | 225 | 11.69 | 0.95 | |
| | 284.12 - 557.52 | 275 | 14.29 | 1.16 | |
| | 557.52 - 2062.83 | 0 | 0.00 | 0.00 | |
| | | | | | |
| Rainfall | 267 - 287 | 1575 | 81.82 | 4.16 | 4.34 |
| | 288 - 299 | 300 | 15.58 | 0.77 | |
| | 300-315 | 0 | 0.00 | 0.00 | |
| | 316-330 | 0 | 0.00 | 0.00 | |
| | 331-364 | 50 | 2.60 | 0.13 | |
| | | | | | |
| Elevation | 0 - 3.231 | 1075 | 55.84 | 3.59 | 3.42 |
| | 3.231 - 16.155 | 825 | 42.86 | 1.90 | |
| | 16.155 - 80.777 | 25 | 1.30 | 0.07 | |
| | 80.777 - 258.486 | 0 | 0.00 | 0.00 | |
| | 258.486 - 481.431 | 0 | 0.00 | 0.00 | |
| | 481.431 - 823.925 | 0 | 0.00 | 0.00 | |
| | | | | | |
| LULC | Built up | 1675 | 87.01 | 1.96 | 3.19 |
| | Low Vege | 250 | 12.99 | 1.29 | |
| | Agriculture | 0 | 0.00 | 0.00 | |
| | Water | 0 | 0.00 | 0.00 | |
| | Forest | 0 | 0.00 | 0.00 | |

ENSEMBLE FREQUENCY RATIO (FR) AND ANALYTICAL
HIERARCHY PROCESS (AHP)

The ensemble between FR and AHP method has been utilized to analyze the ability of this ensemble method in developing flash flood susceptibility mapping of the study area. This ensemble method is using weight from FR analysis and apply as the input data for AHP method. FR values were normalized (0-1) to perform this ensemble method. Equation 10 was used for obtaining the final map in GIS.

$$FSM_{FR-AHP} = FR_{slope}W_{AHP} + FR_{asp}W_{AHP} \quad (10)$$

$$+ FR_{TWI}W_{AHP} + FR_{SPI}W_{AHP} + FR_{NDVI}W_{AHP}$$

$$+ FR_{river}W_{AHP} + FR_{rain}W_{AHP} +$$

$$+ FR_{DEM}W_{AHP} + FR_{LULC}W_{AHP}$$

The value of CR for this ensemble method is 0 which is this model can be used for analysis. The fractional weights of SPI, Elevation, LULC, Rainfall, Curvature and Slope with weights of 0.131, 0.118, 0.110, 0.150, 0.127, and 0.105, respectively, were found to have significant impact on flash flood occurrences in this study area. The other conditioning factors such as distance from river, aspect, and NDVI have not given significant impact for this study area. Table 3 show the factor weights of the conditioning factors.

TABLE 3. The factor weights of the conditioning factors

| Conditioning factor | Factor weight |
|---|---|
| SPI | 0.131 |
| Elevation | 0.118 |
| LULC | 0.110 |
| TWI | 0.095 |
| River | 0.048 |
| Aspect | 0.035 |
| NDVI | 0.082 |
| Slope | 0.105 |
| Curvature | 0.127 |
| Rainfall | 0.150 |

TABLE 4. Matrix of pair-wise ratings of relative importance of predictors

|  | SPI | DEM | LULC | TWI | River | Aspect | NDVI | Slope | Curvature | Rainfall |
|---|---|---|---|---|---|---|---|---|---|---|
| SPI | 1.000 | 1.107 | 1.184 | 1.374 | 2.731 | 3.781 | 1.601 | 1.248 | 1.026 | 0.869 |
| DEM | 0.903 | 1.000 | 1.070 | 1.241 | 2.467 | 3.416 | 1.446 | 1.127 | 0.927 | 0.785 |
| LULC | 0.845 | 0.935 | 1.000 | 1.160 | 2.307 | 3.193 | 1.352 | 1.054 | 0.867 | 0.734 |
| TWI | 0.728 | 0.806 | 0.862 | 1.000 | 1.988 | 2.752 | 1.165 | 0.908 | 0.747 | 0.633 |
| River | 0.366 | 0.405 | 0.434 | 0.503 | 1.000 | 1.384 | 0.586 | 0.457 | 0.376 | 0.318 |
| Aspect | 0.264 | 0.293 | 0.313 | 0.363 | 0.722 | 1.000 | 0.423 | 0.330 | 0.271 | 0.230 |
| NDVI | 0.625 | 0.692 | 0.740 | 0.858 | 1.706 | 2.362 | 1.000 | 0.780 | 0.641 | 0.543 |
| Slope | 0.801 | 0.887 | 0.949 | 1.101 | 2.189 | 3.030 | 1.283 | 1.000 | 0.823 | 0.697 |
| Curvature | 0.974 | 1.078 | 1.154 | 1.338 | 2.661 | 3.683 | 1.559 | 1.216 | 1.000 | 0.847 |
| Rainfall | 1.150 | 1.273 | 1.362 | 1.580 | 3.142 | 4.350 | 1.841 | 1.435 | 1.181 | 1.000 |

TABLE 5. Estimated eigenvectors of the pairwise rating matrix and weights of predictors of susceptibility

| | SPI | Elev | LULC | TWI | River | Aspect | NDVI | Slope | Curv | R.fall | sum | Fractional Weight | Integer Weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SPI | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 1.306 | 0.131 | 3.781 |
| Elevation | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 0.118 | 1.180 | 0.118 | 3.416 |
| LULC | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 0.110 | 1.103 | 0.110 | 3.193 |
| TWI | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.095 | 0.951 | 0.095 | 2.752 |
| River | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.048 | 0.478 | 0.048 | 1.384 |
| Aspect | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.345 | 0.035 | 1.000 |
| NDVI | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.082 | 0.816 | 0.082 | 2.362 |
| Slope | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 0.105 | 1.047 | 0.105 | 3.030 |
| Curvature | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 0.127 | 1.272 | 0.127 | 3.683 |
| Rainfall | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 0.150 | 1.502 | 0.150 | 4.350 |

## MODEL VALIDATION USING RECEIVER OPERATING CHARACTERISTIC (ROC)

For model validation, this research is using the receiver operating characteristic (ROC) method which is the most popular method to evaluate the accuracy of the result of susceptibility mapping. The ROC method has been used widely by researchers in predictive mapping (Althuwaynee et al. 2014; Khosravi et al. 2016a; Lee & Pradhan 2007; Tehrany et al. 2019b). For training data, 70% of 110 flash flood locations were used to analyze the Success Rate of ROC. The other 30% of flash flood locations were utilized to analyze the Prediction Rate of ROC. The success rate curve was used to analyze the ability of the model and the prediction rate shows the ability of model to develop the flash flood mapping (Khosravi et al. 2016b). To analyze the AUC values, (11) can be utilized.

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \tag{11}$$

where TP (true positive) represent the correctly classified flash flood pixels, TN (true negative) represents the correctly classified non- flash flood pixels, P is the total number of flash flood pixels and N is the total number of non-flash flood pixels.

Costache et al. (2020) mentioned that the Area Under Curve (AUC) is the main statistical method of ROC Curve. AUC method is utilized to acquire the accuracy and the performance of the method used (Bradley 1997). To acquire the AUC of training and validation, the graphs of cumulative percentage of flash flood against the flash flood susceptibility index rank percentage were plotted. The accuracy of model is better when the AUC value is near to 1 (Costache et al. 2020). Hong et al. (2018) stated that the AUC 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). Figure 5 shows the ROC curve of FR and FR-AHP method.

The quantile technique was used in developing flash flood susceptibility mapping index. This map was classified into five susceptible zones which are very low, low, moderate, high and very high. Based on the results, the FR and FR-AHP has almost similar results which is about 90% (FR, FR-AHP = 0.90) in success rate accuracy. For the prediction rate, the FR-AHP method is slightly better in accuracy which is 88.33% (AUC = 0.883) and FR method is 85.62% (AUC = 0.856). Figure 6 shows the flash flood susceptibility mapping using FR and FR-AHP, respectively.
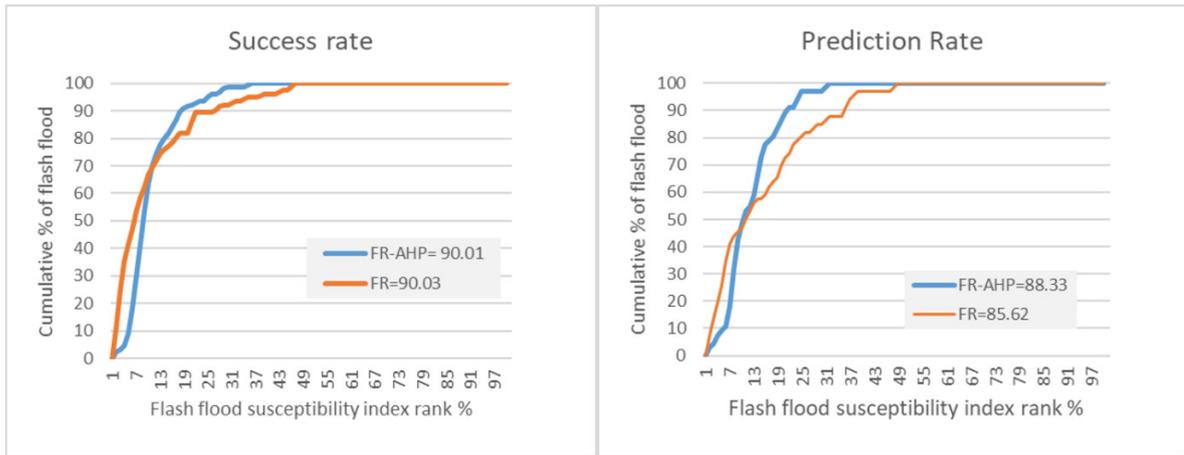
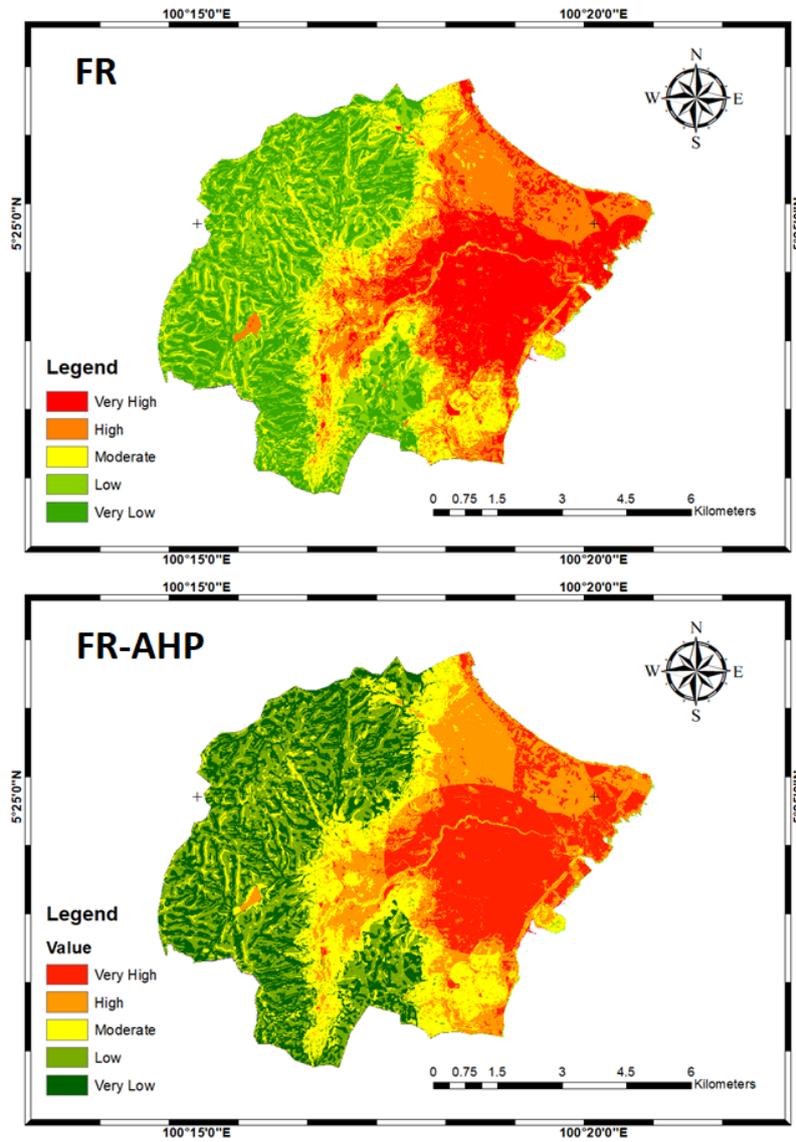FIGURE 5. The ROC curve of FR and FR-AHP method



FIGURE 6. Flash flood susceptibility mapping using FR and FR-AHP

CONCLUSION

Frequency Ratio is one of the popular methods of bivariate statistical analysis (BSA) to develop the predictive mapping. As mentioned before, the frequency ratio method is straightforward, the input and output of this model can be easily understood and able to develop good flash flood susceptibility mapping. This research objective was to analyze the capability ensemble FR with AHP method and make comparison between FR method and FR-AHP method. Another objective of this research was comparing the performance of these two methods on developing flash flood susceptibility map. Ten flash flood conditioning factors (independent) were selected and a flash flood inventory map was used to create the training and validation (dependent) of mapping.

For this study area, the areas with lowest elevation, 0° - 1° or lowest slope angle, convex area, and urban areas are most susceptible areas to flash flooding based on the FR analysis. Quantile method was utilized in maps classification. Based on the result, both methods which are the standalone FR and FR-AHP method get almost similar in success rate, 90.03% and 90.01%, respectively. This good result in success rate which is about 90% can be categorized as very good accuracy. In prediction rate, the FR-AHP method has slightly better accuracy compared to FR method which 88.33% ( FR-AHP) and 85.62% (FR). The output of this study can help the local authority to planning the land use and drainage system. The proper development can also be achieved by identify the location of high risk of flash flood. Therefore, the output of this research is able to assist the local authority in land use planning, drainage system and reduce the impact of flash flood of study area. For the future work, it is recommended to develop the flash flood susceptibility mapping using machine learning method due to capability machine learning to analyze the large and complex data.

REFERENCES

Althuwaynee, O.F., Pradhan, B., Park, H.J. & Lee, J.H. 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena* 114: 21-36.

Bradley, A.P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30(7): 1145-1159.

Bui, D.T., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J.F., Melesse, A.M., Thai Pham, B., Pourghasemi, H.R., Mahmoudi, M., Bahrami, S. & Pradhan, B. 2019a. Flood spatial modeling in Northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sensing* 11(13): 1589.

Bui, D.T., Ngo, P.T., Pham, T.D., Jaafari, A., Minh, N.Q., Hoa, P.V. & Samui, P. 2019b. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena* 179: 184-196.

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L. & Niu, C. 2016. Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. *Sustainability* 8(9): 948.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T. & Khosravi, K. 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling & Software* 95: 229-245.

Chee, H.L., Tan, D.T., Chan, N.W. & Zakaria, N.A. 2018. Applying a system thinking approach to explore root cause of river pollution: A preliminary study of Pinang river in Penang state, Malaysia. In *Proceedings of the 21st IAHR-APD Congress 2018.* pp. 1441-1448.

Conforti, M., Aucelli, P.P.C., Robustelli, G. & Scarciglia, F. 2011. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Natural Hazards* 56(3): 881-898.

Costache, R. 2019. Flash-flood potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models. *Science of the Total Environment* 659: 1115-1134.

Costache, R. & Bui, D.T. 2019. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Science of the Total Environment* 691: 1098-1118.

Costache, R., Popa, M.C., Bui, D.T., Diaconu, D.C., Ciubotaru, N., Minea, G. & Pham, Q.B. 2020. Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *Journal of Hydrology* 585: 124808.

Dano, U.L., Balogun, A.L., Matori, A.N., Wan Yusouf, K., Abubakar, I.R., Said Mohamed, M.A., Aina, Y.A. & Pradhan, B. 2019. Flood susceptibility mapping using GIS-based analytic network process: A case study of Perlis, Malaysia. *Water* 11(3): 615.

Ghosh, S., John, E., Carranza, M., Westen, C.J.V., Jetten, V.G. & Bhattacharya, D.N. 2011. Geomorphology Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). *Geomorphology* 131(1-2): 35-56.

Hashim, M., Misbari, S. & Pour, A.B. 2018 Landslide mapping and assessment by integrating Landsat-8, PALSAR-2 and GIS techniques: A case study from Kelantan State, Peninsular Malaysia. *Journal of the Indian Society of Remote Sensing* 46(2): 233-248.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X. & Chen, W. 2018. Application of fuzzy weight of evidence and data

mining techniques in construction of flood susceptibility map of Poyang County, China. *Science of the Total Environment* 625: 575-588.

Hussaini, H.A. 2007. *Flood and Drought Management in Malaysia*. Kuala Lumpur: Ministry of Natural Resources and Environment. Speech given on 21 June 2007.

Jebur, M.N., Pradhan, B., Shafri, H.Z.M.M., Yusoff, Z.M. & Tehrany, M.S. 2015. An integrated user-friendly ArcMAP tool for bivariate statistical modelling in geoscience applications. *Geoscientific Model Development* 8(3): 881-891.

Khosravi, K., Pham, B.T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I. & Bui, D.T. 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment* 627: 744-755.

Khosravi, K., Nohani, E., Maroufinia, E. & Pourghasemi, H.R. 2016a. 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. *Natural Hazards* 83(2): 947-987.

Khosravi, K., Pourghasemi, H.R., Chapi, K. & Bahri, M. 2016b. 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. *Environmental Monitoring and Assessment* 188(12): 1-21.

Kia, M.B., Moradi, A., Pirasteh, S., Pradhan, B., Mahmud, A.R., Sulaiman, W.N.A. & Moradi, A. 2011. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences* 67(1): 251-264.

Koirala, S. 2011. Some flood damage estimates of the world. *UNESCO-IHE Flood Management Education Platform*. http://www.unesco-ihe.org/Flood-Management-Education-Platform Accessed on January 2012.

Lee, S. & Pradhan, B. 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4(1): 33-41.

Li, L., Lan, H., Guo, C., Zhang, Y., Li, Q. & Wu, Y. 2017. A modified frequency ratio method for landslide susceptibility assessment. *Landslides* 14(2): 727-741.

Pradhan, B. & Youssef, A.M. 2011. A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia. *Journal of Flood Risk Management* 4(3): 189-202.

Pradhan, B., Mansor, S., Pirasteh, S. & Buchroithner, M.F. 2011. Landslide hazard and risk analyses at a landslide prone catchment area using statistical based geospatial model. *International Journal of Remote Sensing* 32(14): 4075-4087.

Pradhan, B., Suliman, M.D.H. & Awang, M.A. 2007. Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS). *Disaster Prevention and Management: An International Journal* 16(3): 344-352.

Rahmati, O., Darabi, H., Haghighi, A.T., Stefanidis, S., Kornejady, A., Nalivan, O.A. & Bui, D.T. 2019. Urban flood hazard modeling using self-organizing map neural network. *Water* 11(11): 2370.

Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T. & Akgun, A. 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences* 7(2): 725-742.

Saleh, A., Yuzir, A. & Abustan, I. 2020. Flash flood susceptibility modelling: A review. In *IOP Conference Series: Materials Science and Engineering*. p. 012005.

Samanta, R.K., Bhunia, G.S., Shit, P.K. & Pourghasemi, H.R. 2018. Flood susceptibility mapping using geospatial frequency ratio technique: A case study of Subarnarekha river basin, India. *Modeling Earth Systems and Environment* 4(1): 395-408.

Tehrany, M.S., Jones, S. & Shabani, F. 2019a. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena* 175: 174-192.

Tehrany, M.S., Jones, S., Shabani, F., Martínez-Álvarez, F. & Bui, D.T. 2019b. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theoretical and Applied Climatology* 137(1-2): 637-653.

Tehrany, M.S., Shabani, F., Jebur, M.N., Hong, H., Chen, W. & Xie, X. 2017. GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. *Geomatics, Natural Hazards and Risk* 8(2): 1538-1561.

Tehrany, M.S., Pradhan, B., Mansor, S. & Ahmad, N. 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena* 125: 91-101.

Tehrany, M.S., Lee, M.J., Pradhan, B., Jebur, M.N. & Lee, S. 2014. Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences* 72(10): 4001-4015.

Tehrany, M.S., Pradhan, B. & 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. *Journal of Hydrology* 504: 69-79.

The Star. 2018. Penang population to hit 1.8mil. *The Star Online*.

Tjaardstra, A. 2015. Malaysia flood damage to cost $560m. *Post UK*. Accessed on 5 April 2019.

Youssef, A.M. & Hegab, M.A. 2019. Flood-hazard assessment modeling using multicriteria analysis and GIS: a case study - Ras Gharib area, Egypt. *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, edited by Pourghasemi, H.R. & Gokceoglu, C. Amsterdam: Elsevier. pp. 229-257.

Youssef, A.M., Pradhan, B. & Sefry, S.A. 2016. Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi

Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences* 75(1): 1-16.

Zakaria, S.F., Zin, R.M., Mohamad, I., Balubaid, S. & Mydin, S.H. 2017. The development of flood map in Malaysia. In *AIP Conference Proce*edings. p. 110006.

Azlan Saleh & Ali Yuzir*
Disaster Preparedness & Prevention Centre (DPPC)
Malaysia-Japan International Institute of Technology (MJIIT)
Universiti Teknologi Malaysia (UTM)
54100 Kuala Lumpur, Federal Territory
Malaysia

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

*Corresponding author; email: muhdaliyuzir@utm.my