SPATIAL PREDICTION OF FLOOD IN KUALA LUMPUR CITY OF MALAYSIA USING LOGISTIC REGRESSION

Abdulwaheed Tella[1*], Muhammad Raza Ul Mustafa[1,2*], Abdullateef Oluwagbemiga Balogun[3,4], Chukwuma J. Okolie[5*], Idris Bello Yamusa[6], Muhammad Bello Ibrahim[1]

[1]Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, 32610, Seri Iskandar, Perak Darul Ridzuan, Malaysia – tellaabdulwaheed01@gmail.com, raza.mustafa@utp.edu.my, muhammad_17006885@utp.edu.my

[2]Centre for Urban Resource Sustainability, Institute of Self-Sustainable Building, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia - raza.mustafa@utp.edu.my

[3]Department of Computer and Information Science, Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Perak, Malaysia - bharlow058@gmail.com

[4]Department of Computer Science, University of Ilorin, Ilorin 1515, Nigeria - bharlow058@gmail.com

[5]Department of Surveying and Geoinformatics, Faculty of Engineering, University of Lagos, Nigeria - cjohnokolie@gmail.com

[6]Geosciences Department, Universiti Teknologi PETRONAS (UTP), Persiaran UTP, Seri Iskandar 32610, Perak, Malaysia - idrisbyamusa@yahoo.com

***Corresponding author**

Commission IV, WG 7

KEYWORDS: Urban Flooding, Logistic Regression, Malaysia, Flood Susceptibility, Flood Hazards

ABSTRACT:

Flooding is one of the most prevalent natural disasters affecting people worldwide. Flooding is a devastating natural disaster in Malaysia regarding the number of people affected, socioeconomic damage, severity, and scale of the impact. Urban flooding is currently a major concern due to the possible consequences and frequency with which it occurs in urban areas as urbanization and population increase. Due to the paved surfaces, paved roads, high population, and buildings that prevent water infiltration and movement to the nearby river, urban floods pose a significant threat to the sustainability of lives and properties in the city. The recent floods in Kuala Lumpur in December 2021 and January 2022 affected many buildings, infrastructure, and lives. As a result, this city needs to model the susceptibility of flood-prone areas for an early warning system against future flood hazards in Kuala Lumpur. This is because flooding can never be eradicated but can be minimized and managed. Therefore, this study integrates geospatial technology and a statistical model (logistic regression) to assess flood hazards in Kuala Lumpur. Ten flood conditioning factors such as altitude, slope, TWI, drainage density, distance to river, LULC, NDVI, NDWI, rainfall and MNDWI were used to predict the areas susceptible to flood. The prediction shows an overall accuracy of 0.84, precision of 0.91, recall of 0.72, and F1-score of 0.80. Distance to river, MNDWI, TWI, and LULC are the critical variables that showed high significance in the model prediction. Thus, stakeholders should prioritize urban planning and increase the drainage system to avoid flood effects.

1. INTRODUCTION

In recent years, rapid urbanisation and climate change have increased the likelihood of natural disasters, such as flooding and the resulting loss of life and property (Tehrany et al. 2014). Flood is one of the most catastrophic natural disasters that is widely known in the world to have caused harm to people, property, and infrastructure (Bubeck and Thieken 2018). Economic losses from flood disasters have greatly risen in the last decade due to increased urbanisation, which includes the conversion of vegetated areas to buildings. It was reported that over 200 million people were affected, and economic losses of $95 billion were incurred from 2011 to 2012 (Chapi et al. 2017). Therefore, it has been established that floods are more dangerous than other natural hazards such as volcanoes, landslides, and earthquakes (Bolt et al. 2013).

In Asia, especially in areas prone to tropical cyclones like Southeast Asia, flooding is responsible for about 90% of all fatalities (Pham et al. 2020). In Malaysia, floods occur almost every year. Peninsular Malaysia experienced torrential rains for three days in 2021. Eight states on the peninsula were affected by the subsequent floods, which claimed at least 54 lives and left 2 people missing (Bernama 2022). More than 71,000 people were concurrently displaced, and over 125,000 people were ultimately impacted. Such disasters have the potential to cause unimaginable damage (France-Presse 2021). Therefore, flood control and mitigation strategies are highly needed due to the immense and potentially irreversible damage that could be done to urban infrastructures, roads, farmlands, and bridges. For decision-makers and stakeholders to limit the amount of damage caused by floods, early warning systems and emergency responses are required. Flooding can cause harm or even death to people, so preventing them is preferable to making up for the losses. According to Bubeck et al. (2012), flood prediction reduces flood-related fatalities and associated financial losses, and identifying flood-prone areas is essential to any strategy for reducing flood damage (Sarhadi et al. 2012). The efficiency of forecasting and controlling floods can be achieved by combining the emerging modelling methods with remote sensing and Geographical Information System (GIS) technologies (Shafapour Tehrany et al. 2019). Taking the necessary precautions can mitigate the current unchecked adverse effects of flooding on coastal and socio-ecological regions (Novelo-Casanova and Rodríguez-Vangort 2016).

There exist different techniques to map flood-prone areas. The four main categories of the most popular techniques are hydrologically based (Farooq et al. 2019; Loi et al. 2019), quantitative (Arora et al. 2021; Cao et al. 2016), qualitative (Balogun et al. 2022; Tella and Balogun 2020), and machine learning techniques (Khosravi et al. 2019; Rahman et al. 2019). The most widely used methods for mapping flooding include artificial neural networks (ANNs), frequency ratio (FR), logistic regression (LR), decision trees (DT), and support vector machines (SVMs) (Mojaddadi et al. 2017; Tehrany et al. 2015). Despite the existence of models for mapping flood susceptibility, a crucial challenge with flood prediction maps is their reliability and precision. Each technique differs from the others in terms of its capabilities and is susceptible to various sources of uncertainty (Shafapour Tehrany et al. 2019). It is noteworthy that some previous studies have ascertained the accuracy of Logistic Regression (LR) for flood susceptibility mapping in other countries (Pham et al. 2020; Shafapour Tehrany et al. 2019; Tsangaratos and Ilia 2016). Therefore, this study aims to map the flood susceptibility in Kuala Lumpur, a metropolitan and urban city in Malaysia, using Logistic Regression (LR). This study aims to answer (i) which conditioning factors greatly affect flooding in Kuala Lumpur, and (ii) how reliable is LR model for flood susceptibility mapping in Kuala Lumpur.

2. STUDY AREA

Kuala Lumpur is a metropolitan city in Malaysia, with an estimated land area of 243 km^2 (Althuwaynee and Pradhan 2017). It has a relatively flat terrain with an elevation range of 0 to 420m above sea level (Althuwaynee et al. 2020). The study area is also referred to as Kuala Lumpur Extended Mega Urban Region, the centroid for socioeconomic development, coupled with Malaysia's national capital (Halim et al. 2020). Over 33% of Malaysia's total population resides in Kuala Lumpur, with an expectance of increase in the population density in the future (Abdul Samad and Shaharudin 2017). Fig.1 shows the location of the study area.

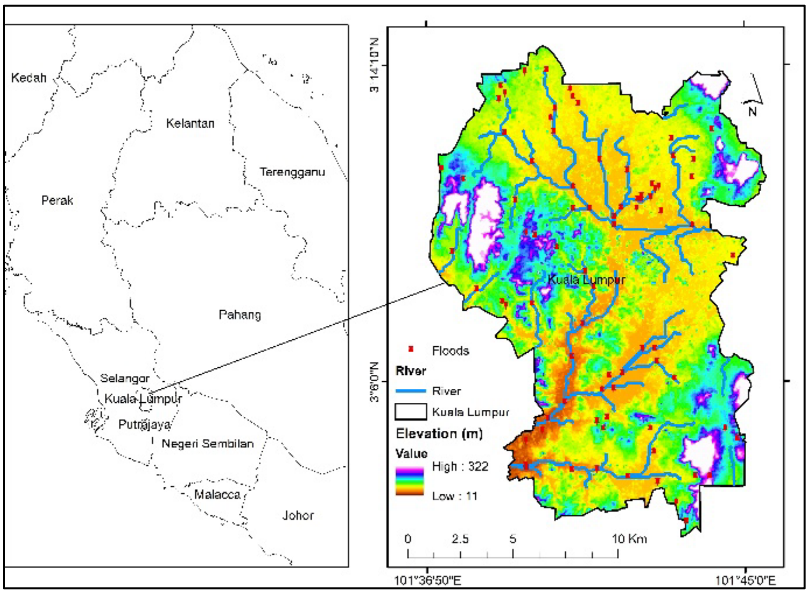

Figure 1: Location of the study area

3. DATA

3.1 Flood Inventory data

The flood inventory is essential for training the LR model and validating the flood susceptibility models. A flood inventory map is necessary to understand the relationship between flood conditioning factors and flood incidences (Tehrany and Jones 2017). To generate susceptibility maps, it is crucial to have a highly accurate record of the previous flooding (Shafapour Tehrany et al. 2019). The primary source of the flood inventory used in this study was documentation of the flooding incident that occurred in the past in Kuala Lumpur, such as the most recent flood events in 2021 and 2022. In addition, Normalized Difference Water Index (NDWI) was used to identify the flood and non-flood points as previously done by Arora et al. (2021), Islam et al. (2022), and Tamiru and Wagari (2022). A total of 82 flood and non-flood points were later identified and used for the training and validation of the LR model. Thus, these data are typically split into training and testing to train the model and validate the results. The flood inventory data were randomly divided into 70% for training and 30% for testing in accordance with previous flood modelling studies (Ali et al. 2020; Pham et al. 2020; Khosravi et al. 2016).

3.2 Flood Conditioning Factors

The most crucial phase in creating the final flood susceptibility maps is the choice of the flood influencing factors, also called conditioning factors, which have a critical effect on the output maps (Shafapour Tehrany et al. 2019). Therefore, the most pertinent and frequently used flood conditioning factors by other researchers (Tehrany et al. 2014; Rahmati et al. 2016; Balogun et al. 2022; Tella and Balogun 2020) were used in this study, even though there is still no universal approach or consensus on how to choose these factors. The complex process of flood susceptibility modelling necessitates the identification of various flood conditioning factors. Therefore, this current study chose ten (10) pertinent factors strongly related to flood hazards based on various studies and experts' opinions. The selected parameters are Altitude, Distance to River, Drainage Density, Land Use/Land Cover (LULC), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Rainfall, Slope,

https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-363-2023 | © Author(s) 2023. CC BY 4.0 License.

and Topographic Wetness Index (TWI) as shown in Figures 2 and 3.

Altitude denotes an essential morphometric factor that has an inverse relationship with flood hazards (Chapi et al. 2017). This means that the higher the altitude, the lower the susceptibility to flood hazards, while low-altitude regions have high flood propensity (Vojtek and Vojteková 2019; Cao et al. 2016). One key determining factor for flood susceptibility mapping is topography, which directly affects the flow and runoff speed (Ali et al. 2020). In addition, the likelihood of flooding is directly influenced by slope through surface runoff and infiltration potential (Chapi et al. 2017). The water infiltration rate decreases and water velocity increases with increasing slope angle. Because of this, flooding is more likely to occur in low-lying areas, such as those near rivers or flat areas (Khosravi et al. 2016). The Topographic Wetness Index (TWI) is an essential hydrologic factor that can be used to assess flood-prone areas (Tien Bui et al. 2016). The topographic wetness index is also referred to as the compound topographic index. The TWI indicates the potential areas where water can get accumulated. A high TWI value represents high potentiality due to a low slope, while a low TWI value indicates the low potentiality of water accumulation (Samanta et al. 2018). TWI is calculated using Equation 1 (Beven and Kirkby 1979):

$$\mathrm{TWI} = In\left(\frac{\mathrm{As}}{tan\beta}\right) \quad (1)$$

Where: A_s = local upslope area ($m^2\ m^{-1}$), $tan\ \beta$ = local slope gradient

According to Tien Bui et al. (2016), the distance to stream and drainage density are two critical factors that affect flooding. The closer to the river, the higher the risk of flood hazards. Also, the drainage density is an essential factor for mapping flood hazards because a higher drainage density increases the surface runoff and the probability of flood hazards. Drainage density (DD) can be calculated by dividing the total length of the rivers in the drainage basin by the total area of the drainage basin. The formula for calculating the DD is shown in Equation 2:

$$Drainage\ Density = \frac{Total\ Stream\ Length}{Area\ of\ Basin} \quad (2)$$

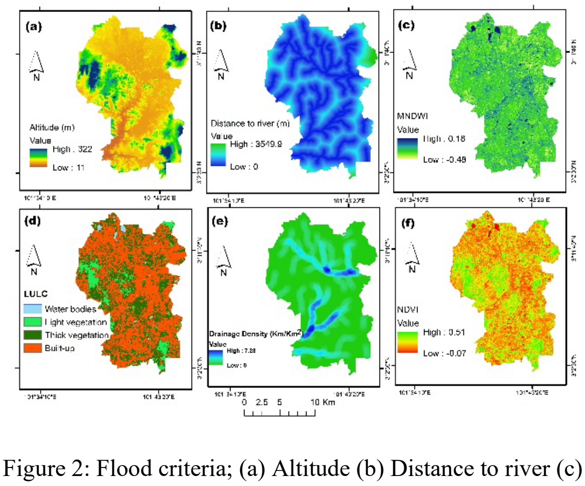

Figure 2: Flood criteria; (a) Altitude (b) Distance to river (c) MNDWI (d) LULC (e) drainage density (f) NDVI

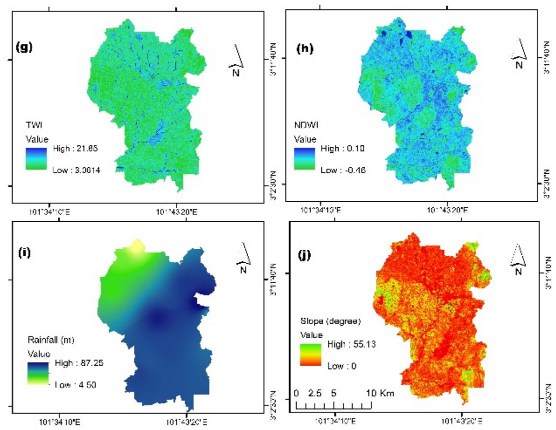

Figure 3: Flood criteria (g) TWI (h) NDWI (i) Rainfall (j) Slope

The Digital Elevation Model (DEM) is one of the most popular datasets for deriving information on topographic elements related to flood susceptibility mapping. The five factors mentioned above are generated from DEM data using the spatial analyst tool in ArcGIS 10.8.

Flooding and vegetation density have an inverse relationship because areas with high vegetation density are less prone to flooding than regions with scanty or no vegetation cover (Khosravi et al. 2019). The Normalized Difference Vegetation Index (NDVI) is an important indicator of flood proneness. NDVI value ranges from +1 to -1, with values of 0.1 or below indicating non-vegetated areas such as soil, bare land, and snow. NDVI index values of 0.2 to 0.3 denote grassland and shrub, while higher values of 0.6 and above represent forest and thick vegetation (Ali et al. 2020). NDVI is calculated from Landsat 8 imagery using Equation 3.

$$\mathrm{NDVI} = \frac{\mathrm{NIR - RED}}{\mathrm{NIR + RED}} \quad (3)$$

Land Use and Land Cover (LULC) significantly affect flood frequency because it regulates surface runoff, flow velocity and sediment transportation (Tien Bui et al. 2020; Costache et al. 2020). Moreover, the water infiltration of surface runoff is significantly impacted by LULC. In addition, built-up areas are greatly susceptible to flooding because buildings reduce water infiltration rates and increase surface water (Samanta et al. 2018). Among the essential factors of flooding is rainfall (Pham et al. 2019; Rahman et al. 2019). The underground hydrostatic level and water pressure are raised by prolonged periods of heavy rainfall (Cao et al. 2016). Furthermore, there is typically a high likelihood of major flood events when there is a lot of rain from the upstream point in a short time. Therefore, flooding is thought to be more likely to occur in areas with high annual rainfall (Chapi et al. 2017). In order to map floods, NDWI is used to identify water bodies or wetlands. Utilizing this index, one can measure how wet or dry a landscape is in relation to moisture (Hakdaoui et al. 2019). The NDWI is extracted from the Near-infrared and green bands (McFeeters 1996), as shown in Equation 4.

$$NDWI = \frac{(Green - NIR)}{(Green + NIR)} \quad (4)$$

MNDWI is a modified version of NDWI (McFeeters 1996). MNDWI can improve the detection of surface water bodies while effectively reducing or eliminating noise from built-up areas, soil and vegetation cover. Equation 5 is used to derive MNDWI.

$$MNDWI = \frac{(Green - SWIR)}{(Green + SWIR)} \quad (5)$$

4. METHODS

4.1 Logistic Regression (LR)

Logistic regression establishes the relationship between the independent variable (X) and the dependent variable (Y) (Kavzoglu et al. 2014; Pradhan and Lee 2010). The dependent variable (training data) is expressed in binary format as 1 and 0, where 1 denotes the presence of a given phenomenon and 0 denotes its absence (Ali et al. 2020). The independent variable can therefore be expressed as discrete data such as elevation, slope, TWI and continuous data such as land cover and lithology. LR predicts spatial susceptibility using the spatial relationship between the dependent and independent variables. In this study, flooded and non-flooded areas were assigned values 1 and 0, respectively. The LR is based on a generalized linear model that can be calculated using Equation 6 (Ali et al. 2020; Tsangaratos and Ilia 2016):

$$P = \frac{1}{(1 + e^{z})} \quad (6)$$

Where P denotes the probability of the flood occurrence. z is a value that ranges from −∞ to +∞ defined by Equation 7.

$$z = b_0 + b_1x_1 + b_2x_2 \cdots + b_nx_n \quad (7)$$

Where b_0 is the intercept of LR model, b_i *(i = 0, 1, 2 … n)* is the slope coefficients and x_i *(i = 0, 1, 2 … n)* are the predictors.

5. RESULTS AND DISCUSSION

Flood hazards are driven by anthropogenic factors, natural influence, and various other predictors associated with geo-environmental features. This study used ten flood conditioning factors: elevation, slope, distance to river, TWI, drainage density, MNDWI, NDWI, LULC and rainfall. The elevation of Kuala Lumpur ranges from 11m to 332m above sea level. The northern regions, northeast and central part, have moderately low elevation. Also, a water body exists around the southern region of Kuala Lumpur. It is also observed that the city has a low vegetation density compared to occupied lands as settlements and built-up areas. This also correlates with the LULC of the study area. Kuala Lumpur is covered with approximately 59% built-up area, 39% light and thick vegetation cover, and 2% water bodies. Since higher drainage density means higher surface runoff and a high probability of flooding, it is observed that areas with roughly 7 km/km^2 correlate with low elevation, meaning that low elevation is highly susceptible to flooding.

Moreover, the MNDWI is better than NDWI for delineating surface water bodies by eliminating some noise in the satellite imagery of built-up areas, vegetation and bare land (Xu 2006). According to Xu (2006), NDWI overestimates the surface water and could not clearly differentiate between water bodies and other land surface features. On the other hand, the TWI clearly indicated the rivers in the study area with high TWI denoting high potentiality of flooding and vice versa. The TWI result also correlates with the distance from river generated for this study. Therefore, four predominant factors, such MNDWI, LULC, distance to river and altitude have a high significance in the LR model. For instance, using ANOVA in R as shown in Table 1 established that MNDWI, LULC, distance to river and altitude have the highest significance on the dependent variable variation, respectively.

Table 1: ANOVA Test result

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			163	227.35	
Altitude	1	6.391	162	220.96	**0.0114725**
Distance to river	1	11.193	161	209.77	**0.0008212**
Drainage distance	1	0	160	209.77	0.9995575
LULC	1	22.687	159	187.08	**1.91E-06**
MNDWI	1	58.525	158	128.56	**2.01E-14**
NDVI	1	0.024	157	128.53	0.8776000
NDWI	1	2.023	156	126.51	0.1549344
Rainfall	1	0.021	155	126.49	0.8838542
Slope	1	0.125	154	126.36	0.7234698
TWI	1	2.968	153	123.4	0.0849227

Bold figures represent the most significants variables

To ascertain how a quantitative predicted variable change based on the levels of one or more predictors, the statistical test known as the analysis of variance (ANOVA) is utilized. An ANOVA is run in R to see if the group means differ between the groups. Thus, it can be concluded that these four criteria are essential factors for mapping flood hazards. Figure 4 shows the flood susceptibility map generated by the LR model.

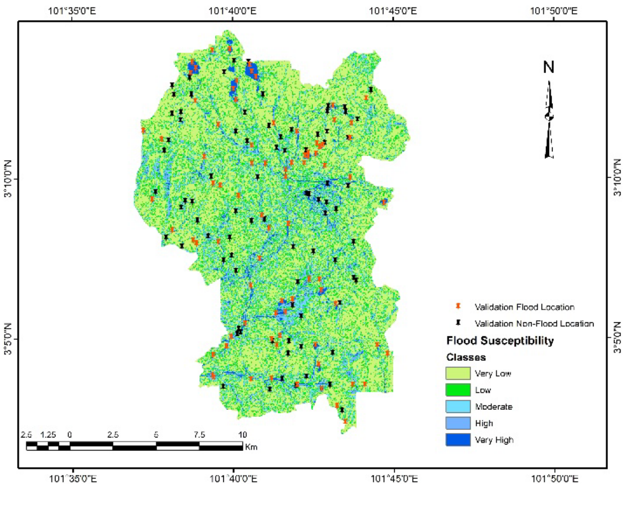

Figure 4: Flood susceptibility map generated by LR model

It is observed from Figure 4 that the water bodies are generally categorized as high to very high susceptibility to flood hazards. The moderate susceptible regions are somewhat close to the water bodies, meaning that these areas or regions can probably be affected by flood hazards if proper management and control are not implemented. Figure 5 shows the area covered by the flood susceptibility classes. It is observed that the areas covered in decreasing order are very low (54.11%), low (23.44%), moderate (12.56%), high (6.60), and very high (3.30%). Although, a high percentage of the study area has low susceptibility to flooding. Nevertheless, factors such as distance to river and LULC that have a high significance in the model should be prioritized for policy formulation and implementation.

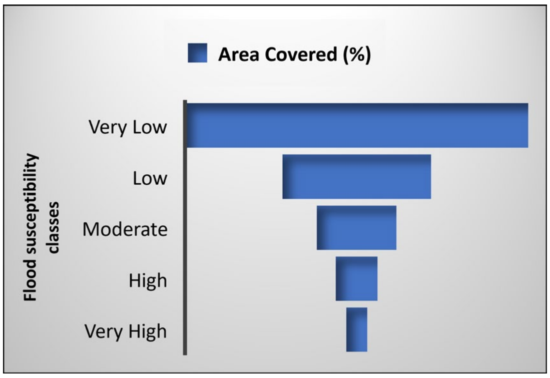

Figure 5: Area covered in percentage

Since Kuala Lumpur city can be easily affected by flooding due to the high building density leading to a low water infiltration rate and high surface runoff, there should be a considerable distance between buildings or settlements from water bodies. Moreover, proper urban planning should be considered with adequate drainage channels with a large capacity to contain surface runoff. Therefore, it is highly advisable to further research on flood susceptibility using other novel models to compare and generate flood maps for the decision-making process. A confusion matrix was used to validate the flood susceptibility map using accuracy, precision, recall and F1-score. The LR model has an accuracy of 0.84, precision of 0.91, recall of 0.72 and F1 score of 0.80. Notably, none of these metrics has a value less than 0.7 showing the acceptability and applicability of the LR model for flood susceptibility mapping. Also, the ROC curve is shown in Figure 6.

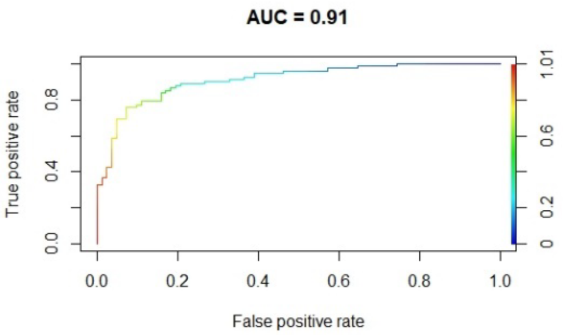

Figure 6: ROC curve of the LR model

6. CONCLUSION

In this study, Logistic Regression (LR) was used to map flood susceptibility in Kuala Lumpur, Malaysia. The model was built using ten (10) geospatial flood conditioning factors as independent variables, with 164 flood events as the dependent variable. Four of the conditioning factors, that is, distance to river, MNDWI, TWI and LULC showed a high significance in the model prediction. The water bodies such as rivers or stagnant water are the primary causes of floods in Kuala Lumpur. Coupled with the high percentage of buildings and cemented surfaces, runoff is bound to occur. Moreover, this can lead to low water infiltration leading to increased flood depth that can put the citizens, infrastructures and motor cars in danger of being flooded. Confusion matrix and AUC or ROC were used to validate the model result. The model has an overall accuracy of 0.84 and model precision of 0.91. The AUC of ROC also established the accuracy of the model used. Continuous research on flood mitigation using different and advanced techniques is therefore recommended for subsequent studies.

References

Abdul Samad, H., and I. Shaharudin. 2017. 'The Emerging Kuala Lumpur Extended Mega Urban Region (KLEMUR): Implications on Urban Prosperity in Malaysia', *Int. J. Malay World Civilis*, 5: 67-74.

Ali, S. A., F. Parvin, Q. B. Pham, M. Vojtek, J. Vojteková, R. Costache, N. T. T. Linh, H. Q. Nguyen, A. Ahmad, and M. A. Ghorbani. 2020. 'GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topľa basin, Slovakia', *Ecological Indicators*, 117: 106620.

Althuwaynee, O. F., A. L. Balogun, and W. Al Madhoun. 2020. 'Air pollution hazard assessment using decision tree algorithms and bivariate probability cluster polar function: evaluating inter-correlation clusters of PM10 and other air pollutants', *GIScience & Remote Sensing*, 57: 207-26.

Althuwaynee, O. F., and B. Pradhan. 2017. 'Semi-quantitative landslide risk assessment using GIS-based exposure analysis in Kuala Lumpur City', *Geomatics, Natural Hazards and Risk*, 8: 706-32.

Arora, A., M. Pandey, M. A. Siddiqui, H. Hong, and V. N. Mishra. 2021. 'Spatial flood susceptibility prediction in Middle Ganga Plain: comparison of frequency ratio and Shannon's entropy models', *Geocarto International*, 36: 2085-116.

Balogun, A.-L., T. Y. Sheng, M. H. Sallehuddin, Y. A. Aina, U. L. Dano, B. Pradhan, S. Yekeen, and A. Tella. 2022. 'Assessment of data mining, multi-criteria decision making and fuzzy-computing techniques for spatial flood susceptibility mapping: a comparative study', *Geocarto International*: 1-27.

Bernama. 2022. 'Sabah Floods: Body of fisherman found, another victim still missing'. https://www.bernama.com/en/general/news_disaster.php?id=2039730.

Beven, K. J., and M. J. Kirkby. 1979. 'A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant', *Hydrological Sciences Bulletin*, 24: 43-69.

Bolt, B. A., W. Horn, G. A. MacDonald, and R. Scott. 2013. *Geological Hazards: Earthquakes-tsunamis-volcanoes-avalanches-landslides-floods* (Springer Science & Business Media).

Bubeck, P., W. J. W. Botzen, and J. C. Aerts. 2012. 'A review of risk perceptions and other factors that influence flood mitigation behavior', *Risk Analysis: An International Journal*, 32: 1481-95.

Bubeck, P., and A. H. Thieken. 2018. 'What helps people recover from floods? Insights from a survey among flood-affected residents in Germany', *Regional environmental change*, 18: 287-96.

Cao, C., P. Xu, Y. Wang, J. Chen, L. Zheng, and C. Niu. 2016. 'Flash Flood Hazard Susceptibility Mapping Using Frequency Ratio and Statistical Index Methods in Coalmine Subsidence Areas', *Sustainability*, 8.

Chapi, K., V. P. Singh, A. Shirzadi, H. Shahabi, D. T. Bui, B. T. Pham, and K. Khosravi. 2017. 'A novel hybrid artificial intelligence approach for flood susceptibility assessment', *Environmental Modelling and Software*, 95: 229-45.

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

Farooq, M., M. Shafique, and M. S. Khattak. 2019. 'Flood hazard assessment and mapping of River Swat using HEC-RAS 2D model and high-resolution 12-m TanDEM-X DEM (WorldDEM)', *Natural Hazards*, 97: 477-92.

France-Presse, A. 2021. 'Malaysia’s worst flooding in years leaves 30,000 people displaced', The Guardian. https://www.theguardian.com/world/2021/dec/19/malaysias-worst-flooding-in-years-leaves-30000-people-displaced.

Hakdaoui, S., A. Emran, B. Pradhan, C.-W. Lee, and S. C. Nguemhe Fils. 2019. 'A Collaborative Change Detection Approach on Multi-Sensor Spatial Imagery for Desert Wetland Monitoring after a Flash Flood in Southern Morocco', *Remote Sensing*, 11.

Halim, N. D. A., M. T. Latif, A. F. Mohamed, K. N. A. Maulud, S. Idrus, A. Azhari, M. Othman, and N. M. Sofwan. 2020. 'Spatial assessment of land use impact on air quality in mega urban regions, Malaysia', *Sustainable Cities and Society*, 63: 102436.

Islam, M. M., K. Ujiie, R. Noguchi, and T. Ahamed. 2022. 'Flash flood-induced vulnerability and need assessment of wetlands using remote sensing, GIS, and econometric models', *Remote Sensing Applications: Society and Environment*, 25: 100692.

Kavzoglu, T., E. K. Sahin, and I. Colkesen. 2014. 'Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression', *Landslides*, 11: 425-39.

Khosravi, K., E. Nohani, E. Maroufinia, and H. R. Pourghasemi. 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', *Natural Hazards*, 83: 947-87.

Khosravi, K., H. Shahabi, B. T. Pham, J. Adamowski, A. Shirzadi, B. Pradhan, J. Dou, H.-B. Ly, G. Gróf, H. L. Ho, H. Hong, K. Chapi, and I. Prakash. 2019. 'A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods', *Journal of Hydrology*, 573: 311-23.

Loi, N. K., N. D. Liem, L. H. Tu, N. T. Hong, C. D. Truong, V. N. Q. Tram, T. T. Nhat, T. N. Anh, and J. Jeong. 2019. 'Automated procedure of real-time flood forecasting in Vu Gia–Thu Bon river basin, Vietnam by integrating SWAT and HEC-RAS models', *Journal of Water and Climate Change*, 10: 535-45.

McFeeters, S. K. 1996. 'The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features', *International Journal of Remote Sensing*, 17: 1425-32.

Mojaddadi, H., B. Pradhan, H. Nampak, N. Ahmad, and A. H. b. Ghazali. 2017. 'Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS', *Geomatics, Natural Hazards and Risk*, 8: 1080-102.

Novelo-Casanova, D. A., and F. Rodríguez-Vangort. 2016. 'Flood risk assessment. Case of study: Motozintla de Mendoza, Chiapas, Mexico', *Geomatics, Natural Hazards and Risk*, 7: 1538-56.

Pham, B. T., T. V. Phong, H. D. Nguyen, C. Qi, N. Al-Ansari, A. Amini, L. S. Ho, T. T. Tuyen, H. P. Yen, H.-B. Ly, I. Prakash, and D. Tien Bui. 2020. 'A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping', *Water*, 12.

Pham, Q. B., S. I. Abba, A. G. Usman, N. T. T. Linh, V. Gupta, A. Malik, R. Costache, N. D. Vo, and D. Q. Tri. 2019. 'Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall', *Water Resources Management*, 33: 5067-87.

Pradhan, B., and S. Lee. 2010. 'Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling', *Environmental Modelling & Software*, 25: 747-59.

Rahman, M., C. Ningsheng, M. M. Islam, A. Dewan, J. Iqbal, R. M. A. Washakh, and T. Shufeng. 2019. 'Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis', *Earth Systems and Environment*, 3: 585-601.

Rahmati, O., H. R. Pourghasemi, and H. Zeinivand. 2016. 'Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran', *Geocarto International*, 31: 42-70.

Samanta, S., D. K. Pal, and B. Palsamanta. 2018. 'Flood susceptibility analysis through remote sensing, GIS and frequency ratio model', *Applied Water Science*, 8: 66.

Sarhadi, A., S. Soltani, and R. Modarres. 2012. 'Probabilistic flood inundation mapping of ungauged rivers: Linking GIS techniques and frequency analysis', *Journal of Hydrology*, 458: 68-86.

Shafapour Tehrany, M., L. Kumar, M. Neamah Jebur, and F. Shabani. 2019. 'Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods', *Geomatics, Natural Hazards and Risk*, 10: 79-101.

Tamiru, H., and M. Wagari. 2022. 'Machine-learning and HEC-RAS integrated models for flood inundation mapping in Baro River Basin, Ethiopia', *Modeling Earth Systems and Environment*, 8: 2291-303.

Tehrany, M. S., and S. Jones. 2017. 'Evaluating the variations in the flood susceptibility maps accuracies due to the alterations in the type and extend of the flood inventory', *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*: 209-14.

Tehrany, M. S., M.-J. Lee, B. Pradhan, M. N. Jebur, and S. Lee. 2014. 'Flood susceptibility mapping using integrated bivariate and multivariate statistical models', *Environmental Earth Sciences*, 72: 4001-15.

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

Tella, A., and A.-L. Balogun. 2020. 'Ensemble fuzzy MCDM for spatial assessment of flood susceptibility in Ibadan, Nigeria', *Natural Hazards*, 104: 2277-306.

Tien Bui, D., N.-D. Hoang, F. Martínez-Álvarez, P.-T. T. Ngo, P. V. Hoa, T. D. Pham, P. Samui, and R. Costache. 2020. 'A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area', *Science of The Total Environment*, 701: 134413.

Tien Bui, D., B. Pradhan, H. Nampak, Q.-T. Bui, Q.-A. Tran, and Q.-P. Nguyen. 2016. 'Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS', *Journal of Hydrology*, 540: 317-30.

Tsangaratos, P., and I. Ilia. 2016. 'Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size', *CATENA*, 145: 164-79.

Vojtek, M., and J. Vojteková. 2019. 'Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process', *Water*, 11.

Xu, H. 2006. 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *International Journal of Remote Sensing*, 27: 3025-33.