Arabian Journal of Geosciences (2018) 11: 765
https://doi.org/10.1007/s12517-018-4095-0

ORIGINAL PAPER

# Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors

Ahmed E. M. Al-Juaidi[1] · Ayman M. Nassar[2] · Omar E. M. Al-Juaidi[3]

Received: 25 May 2018 / Accepted: 16 November 2018 / Published online: 12 December 2018


**Abstract**
This paper investigates the application of logistic regression model for flood susceptibility mapping in southern Gaza Strip areas. At first, flood inventory maps were identified using Palestinian Water Authorities data and extensive field surveys. A total of 140 flood locations were identified, of which 70% were randomly used for data training and the remaining 30% were used for data validation. In this investigation, six causing flood variables from the spatial database were prepared, which are digital elevation model (DEM), topographic slope, flow accumulation, rainfall, land use/land cover (LULC), and soil type. Then, comprehensive statistical analysis techniques including Pearson's correlation, multicollinearity, and heteroscedasticity analyses were used, to ensure that the regression assumptions are not violated. The uniqueness of the current study is its inclusiveness of influential causing flood parameters and vigorous statistical analyses that led to accurate flood prediction. Quantitatively, the proposed model is robust with very reasonable accuracy. The prediction and success rates are 76 and 81%, respectively. The practical and unique contribution of this investigation is the generation of flood susceptibility map for the region. This is a very useful tool for the decision makers in the Gaza Strip to reduce human harm and infrastructure losses.

**Keywords** Flood susceptibility mapping · Logistic regression · Flood conditioning factors · GIS · Southern Gaza strip

## Introduction

Flood is a major destructive natural hazard that can lead to a considerable harmful impact on both economic and social scales. The existing master plans used for land development of urban areas fail to match unpredictable manifestations like flooding (Hsu et al. 2014). Road expansion and urban development could increase the risk of flood occurrence (Iwalewa et al. 2016). Although the flood avoidance is impossible and inevitable, future floods can be delineated through employing forecasting techniques (Tehrany et al. 2015). Therefore, developing a flood prediction model at temporal and spatial levels becomes necessary, which could help in setting up a flood risk alleviation plan and provide disaster assistance services (Schumann et al. 2014).

Establishing the vulnerable regions to floods through preparing of flood susceptibility maps is an essential tool that can reduce future flood destructions. Therefore, identifying locations with high susceptibility to flooding is essential to reduce future floods. Moreover, specifying areas with low susceptibility to flooding could be helpful for development activities (Sarhadi et al. 2012).

Three phases should be considered to mitigate flood according to Kourgialas and Karatzas (2011): pre-flood estimation, flood prediction, and post-flood actions. Konadu and Fosu (2009) mentioned that flood mitigation can be achieved throughout four steps: forecasting, construction and development, precluding, and damage evaluation. Clearly, the identification of flood-vulnerable areas is essential for a quick response and can provide early warning to reduce the impact of likely future flood events (Kia et al. 2012).

Flood problem becomes significant in the Gaza Strip more than in the past few years ago. In December 2013, a winter storm Alexa hits Gaza Strip causing huge damages to property, municipal infrastructure, and agriculture. Alexa is

✉ Ahmed E. M. Al-Juaidi
aealjuaidi@gmail.com; ahmed.aljuaidi@gmail.com

1 Civil Engineering Department, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

2 Civil and Environmental Engineering Department, Utah State University, Logan, UT, USA

3 Business and Finance Administration Department, University college of Applied Sciences, Gaza, Palestine

described as the worst winter storm to hit the region since 1953. Many people have left their houses and transferred to safer places in order to avoid flood risk. Topography, lack of investment in infrastructure, and the high vulnerability of the local communities cause the middle and southern Gaza Strip to be the most seriously impacted by floods (PWA 2012).

Real-time information and predictions of a flood are a valued tool for decision-makers to mitigate their impacts (Chaney et al. 2015). In addition, most of flood prediction work focuses on studying the factors which feature the properties of local flood spots.

Flood forecasting involves interrelated spatially distributed hydro-meteorological parameters which describe meteorological and hydrological characteristics of semi-arid watersheds (Wagener et al. 2007). Therefore, in order to increase the capability to forecast the occurrence of flood events, additional factors depict that the hydrological characteristics of a watershed have considered topographic slope, flow direction, and flow accumulation which represent the contribution area drains into outlets of a watershed (Wagener et al. 2007).

### Previous work

Different methods were employed to describe flood susceptibility and risk mapping using various stochastic methods with GIS and remote sensing. Pradhan (2010) applied multivariate logistic regression with remote sensing and GIS to obtain flood susceptibility mapping in Malaysia. Tehrany et al. (2014a) combined logistic regression with a bivariate probability to construct flood susceptibility maps of Busan City. Chormanski et al. (2011) developed remote sensing and GIS including water chemistry analysis to create flood mapping in Poland. Tehrany et al. (2014b) used a support vector machine (SVM) and weight-of-evidence (WofE) techniques for flood mapping in Malaysia. Youssef et al. (2016) applied frequency ratio (FR) and linear regression (LR) methods and their combinations for flood susceptibility mapping in Jeddah-Saudi Arabia. Omid Rahmati et al. (2016) used WofE and frequency ratio (FR) methods for flood susceptibility mapping in Iran. Baduna Kocyigit et al. (2017) employed HEC-HMS model to study the impact of the size of the sub-basins of a watershed on the hydrologic parameters and hydrograph of adjacent ungauged basins. Akay et al. (2018) used HEC-HMS System for hydrological modeling in the Western Black Sea Region that regularly experiences flooding. Flood susceptibility mapping using hydrological and stochastic rainfall analyses has been used for multiple case studies (Ohlmacher and Davis 2003; Lee and Pradhan 2007).

In this paper, the logistic regression model along with a spatial database of influential flood causing parameters is used to produce the flood susceptibility mapping to Gaza Strip Southern Governorates. The study took into consideration the most significant variables that influence the local characteristics of spots where the flood happened. The variables are the digital elevation model (DEM), soil type, land use/land cover, topographic slope, and flow accumulation. The use of these variables to quantify flood associated attributes is a significant contribution to the current work. Another objective of this study is to find which flood causing variable has the utmost impact in a flood. The novelty of this paper is its inclusiveness of influential flood causing variables on flood occurrence. Another significant outcome of this study is the ability to predict the factors that impact flood occurrence the most. The work employs various statistical analyses (e.g., Pearson correlation, multicollinearity, and heteroscedasticity analyses) to ensure the accuracy of flood predictions. The proposed introduces a very practical tool, first of its kind for the region, for accurate flood predictions.

## Study area

Gaza Strip is situated on the eastern coast of the Mediterranean Sea, and border Egypt on the south and Israel on the North and West, with a total area of 365 km$^2$. The length of the Gaza Strip is approximately 41 km long. The width of the Gaza Strip varies between 6 and 12 km. Nowadays, the population of the Gaza Strip is approximately 1.95 million with a population density of 5342 people/km$^2$ (Al-Juaidi et al. 2014; Al-Juaidi 2018).

Gaza Strip experiences severe water problems. Groundwater aquifer is the major water supply source, used for all water use sectors. This groundwater is recharged only by rainfall, with around 55 million cubic meters per year, and average rainfall of 320 mm per year (Al-Juaidi et al. 2010, 2011; Fisher et al. 2005). The study area contains 16 sub-watersheds. Around 155 million cubic meters of water abstracted from the aquifer with approximately 105 million cubic meters is recharged into the aquifer with shortage equal to 50 million cubic meters (Al-Juaidi et al. 2009; Fisher et al. 2005).

## Methodology

The main steps of this study is to delineate the flood susceptibility mapping and consist of five stages: (1) preparing of flood inventory map and flood conditioning factors from spatial database, (2) statistical analyses methods, (3) logistic regression analysis, (4) generation of the flood susceptibility map, and (5) validation of the results using under the curve (AUC) of the receiver operating characteristics (ROC) method. ArcGIS 10.2 and EViews 10 statistical software were used for data analyses throughout all stages (see Fig. 1).

The main assumption using the logistic regression is the possibility of a flood occurring in the future. The flood susceptibility

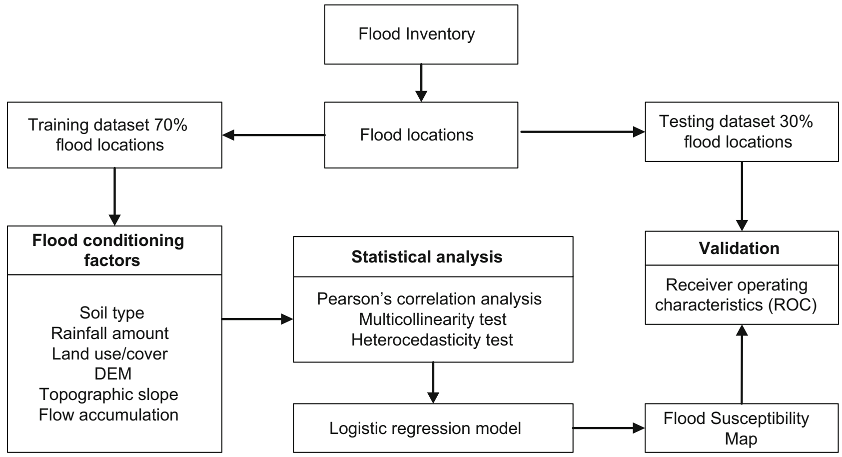



**Fig. 1** Flow chart of the methodology

(probability) mapping obtained from this study will be compared with historical flooding maps. First, flood-causing factors such as land use land cover (LULC), soil type, DEM, and rainfall maps were obtained from the Palestinian Water Authority database. Flow direction and topographic slope were obtained from the Arc-GIS. The flow accumulation map was extracted from flow direction and DEM maps using the spatial analysis tools. Second, various statistical techniques will be used to ensure that the logistic regression parameters are valid and accurate. These techniques include the Pearson correlation, multicollinearity, and heteroscedasticity tests. In the next step, logistic regression was used to find the correlation between the flood occurrence and its reliance on several independent conditioning variables. The fitted parameters obtained from logistic regression were used in the GIS to obtain flood susceptibility map (Eq. 1). Finally, the AUC-ROC method was employed to validate the flood susceptibility map.

## Flood inventory map

One of the essential factors in predicting flood is the inventory map, which corresponds to occurrences of a flood. The flood vulnerable zones have been identified from Palestinian Water Authorities (PWA) reports, extensive field surveying, where the coordinates of these locations have been measured using Global Positioning Systems (GPS). The flood inventory map was prepared by the past floods of 1989, 2003, and 2013 according to PWA extensive field inventories. Total points of 140 flood where floods occurred were surveyed at 1:25,000 scale was sketched by drawing polygons representing these points (Fig. 2). This process has already been assisted and identified by Engineers working in the Gaza Coastal Municipalities and Water Utilities.

To obtain a flood susceptibility map, flood occurrences were classified into two classes, testing, and training. It is recommended that 70% of flood occurrences can be used for training and the rest can be used for validation (Tehrany et al. 2014a). In this study, 98 cases of flood occurrences (70%) were used for training, while the remaining 42 cases of flood occurrences (30%) were used for validation.

## Flood conditioning (causing) factors

Establishing the efficient factors which affect flood occurrence to create flood susceptibility maps is essential (Kia et al. 2012). Therefore, six conditioning factors were created from GIS database. These factors are rainfall amount, land-use/land cover, soil type, DEM, topographic slope, and flow accumulation. These flood causing factors are presented in GIS raster grid with 20 m × 20 m cell size.

## Rainfall amount (mm)

Rainfall is the main resource for groundwater recharge in the Gaza Strip. Surface water is insignificant in the area. Recharge is estimated at 50 million cubic meters per year from rainfall (Fisher et al. 2005). The average annual rainfall in the Gaza Strip is 320 mm/year. The average annual rainfall is 410 mm/year in the north and decreases to 220 mm in the south, as shown in Fig. 3a.

## Land use/land cover (LULC)

Figure 3b shows the distributions of various land use type in the southern Gaza Strip governorates in 2010. The agricultural areas consist of about 60% of the total area. The urban buildup areas represent 20% of the land use, and the natural resource areas approximately cover 20%.

## Soil type

Infiltration of water mainly depends on soil texture. Soil texture is a major factor which has to be fulfilled before the generation of surface run-off. The Gaza Strip has eight

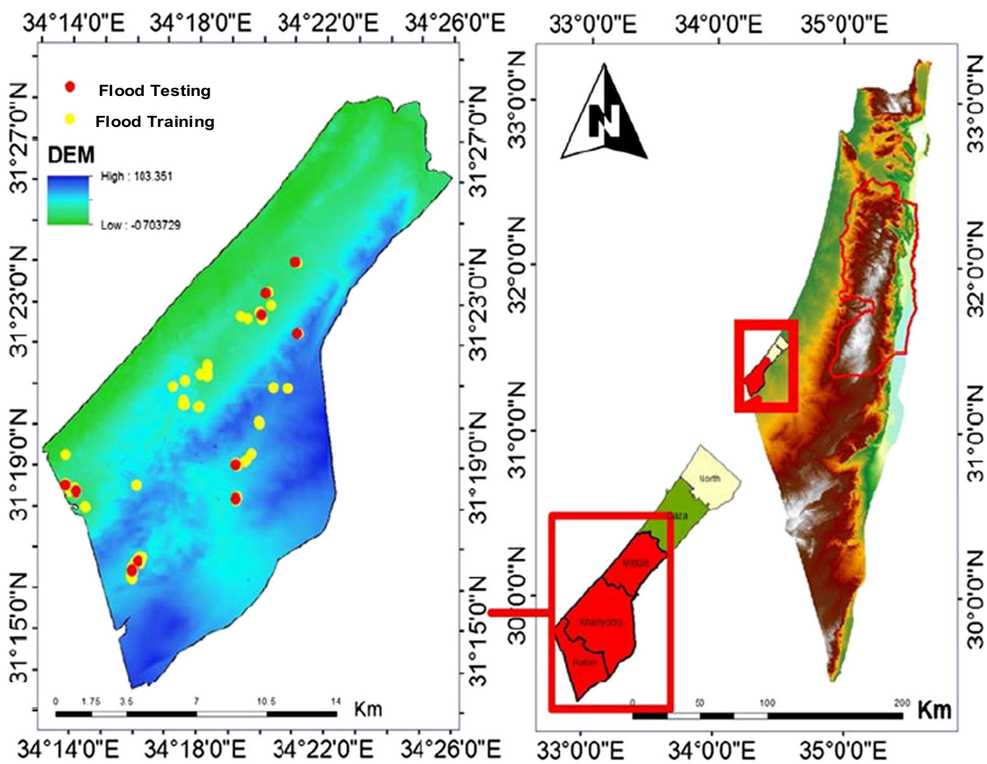
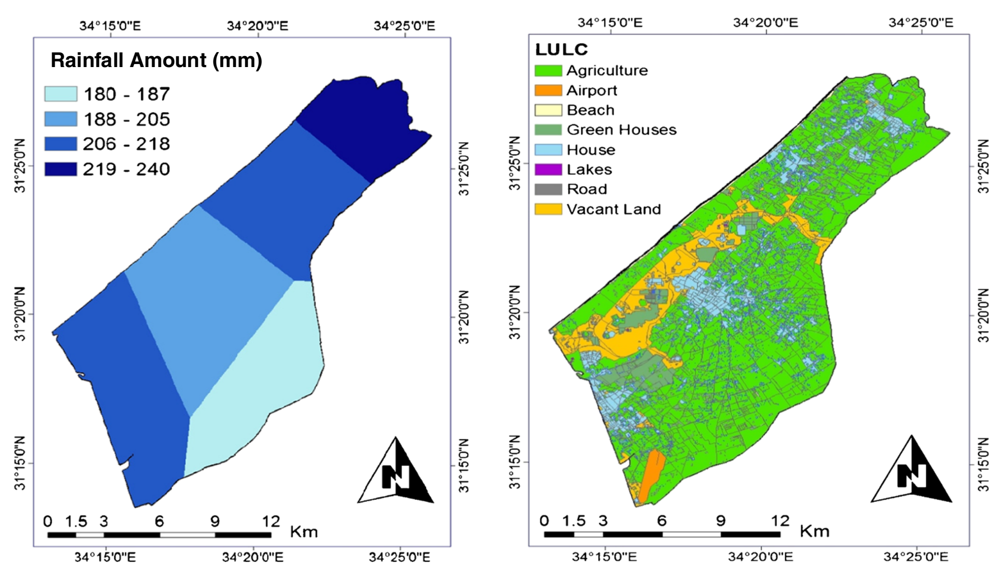



**Fig. 2** Flood inventory map

different soil types as shown in Fig. 4a (PEPA 1996). Sand dunes represent the sandy soil and located along the coastline of the Gaza Strip with a depth range from 2 to about 50 m (Al-Agha 1995). The soil becomes clay, silt, and loess in the east. The soil turns into dark brown (clay) towards the northeast. In valleys, the soil is almost loess type, with a depth of ranges from 25 to 30 m (Aish and De Smedt 2004).

## Digital elevation model (DEM)

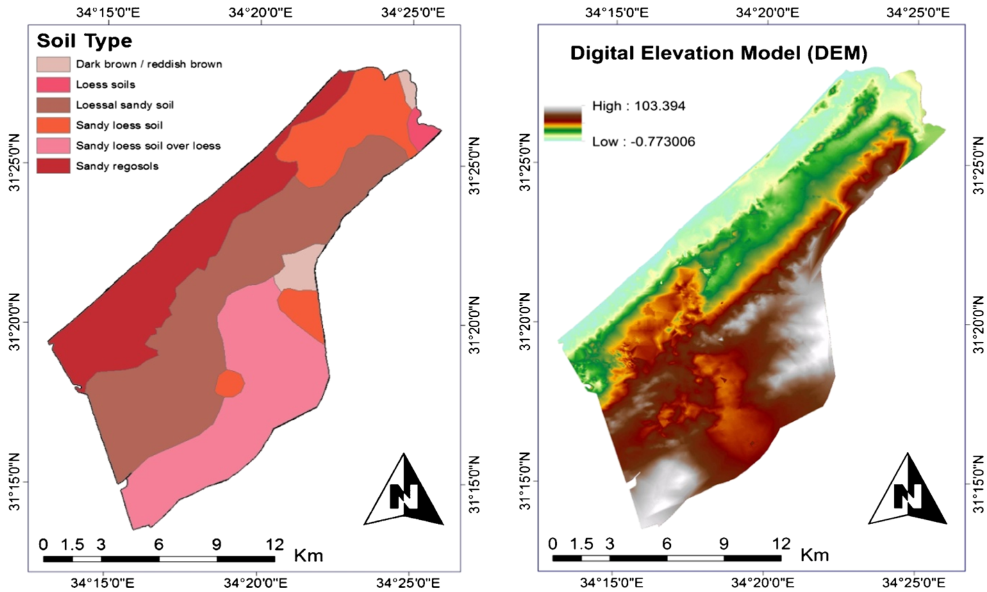
The digital elevation model is one of the important tools for flood prediction, which allows for a three-dimensional vision of the ground surface terrain. The digital elevation model map for the Gaza Strip is obtained from NASA website based on the Automated Geospatial Watershed




a) Rainfall amount map

b) Land use/cover (LULC) map

**Fig. 3** Rainfall and land use/cover (LULC) maps



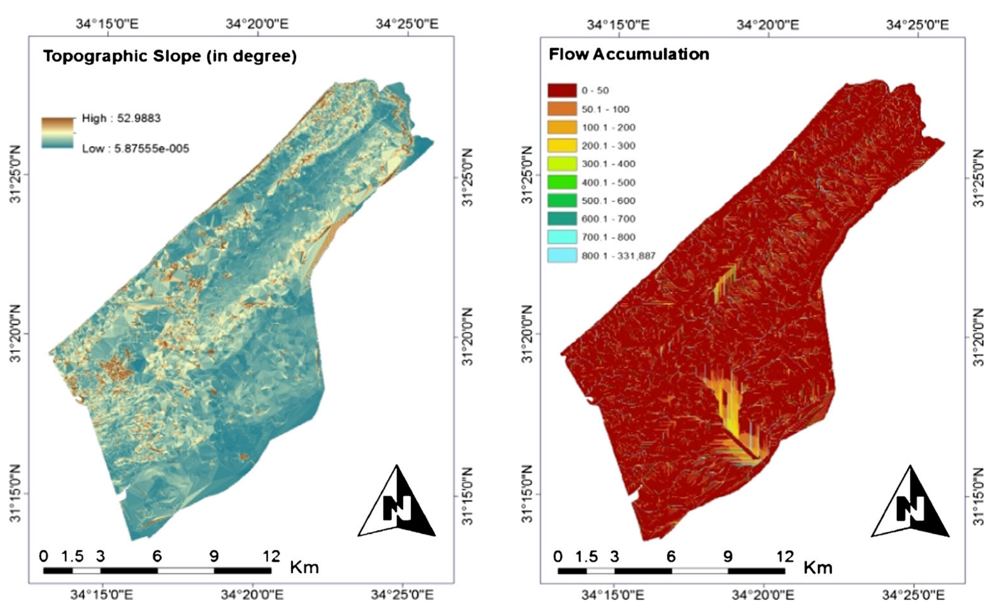

a) Soil type map b) Digital elevation model (DEM) map

**Fig. 4** Soil type and DEM maps

Assessment Tool (AGWA) requirements (see Fig. 4b). Generally, the land topography in Gaza Strip tends to incline from east to west with an irregular shape. The maximum elevation reaches around 110 m above mean sea level at the eastern border while in the west has the same sea level. Most of the stormwater runoff flows from east to west and is collected in the lowest elevation point located at 3-km distance from the sea coastal line.

## Topographic slope (in degree)

Slope controls the velocity of water on the ground surface, which affects infiltration and runoff. In areas where the slope is low, the infiltration depth is high and the runoff is low (Adiat et al. 2012). In order to find out the location of the pool of rainwater, the topographic slope must be considered. The topographic slope map was produced in GIS (see Fig. 5a). In




a) Topographic slope map b) Flow accumulation map

**Fig. 5** Topographic slope and flow accumulation maps order to find out the direction through which water will flow from the cell to the adjacent cells, value for each cell elevation is compared with neighboring cell elevations in GIS.

### Flow accumulation

Flow accumulation is another influencing factor in the flood occurrence (Pradhan 2010). Flow accumulation represents a raster map of accumulated flow to each cell and identified by accumulating the weight of all cells entering into each downslope cell in the GIS raster. Undefined flow direction cells will not contribute to any downstream flow. These cells of undefined flow direction will only receive flow. The accumulated flow depends on the number of cells flowing into each cell in the output raster. Areas of concentrated flow (e.g., high flow accumulation cells) can be used to recognize stream channels. When flow accumulation cell is zero, it depicts local topographic elevations which may be considered to identify watershed ridges.

Flow direction aims to find the direction of the flow (e.g., water and sediments, etc.) which follows the topography. Flow direction is a physically based principle where water flow follows the steepest slope direction. The slope is recognized by determining the plane tangent to the topographic surface in the center of the cell. The slope gradient is the maximum elevation drop within the plane, while the correspondent direction of this elevation change rate represents the aspect. More details about the flow direction methods can be found at Fernando et al. (2008). The flow direction map has been used with the DEM to derive the flow accumulation map using Arc-GIS spatial analysis tools (see Fig. 5b).

## Flood susceptibility modeling

### Logistic regression model

Logistic regression (Hosmer and Lemeshow 2000; Helsel and Hirsch 2002) is applied to predict the probability of an event (e.g., flood) occurring. Logistic regression is similar to multiple linear regressions and permits one to have a relation between flood occurrence (one dependent variable) and flood causing factors (other independent variables) (Helsel and Hirsch 2002). For flood vulnerability analysis, logistic regression aims to obtain the best fitting model to depict the relationship between the dependent and independent variable (Omid Rahmati et al. 2016; Pradhan 2010). In logistic regression, the more conditioning variables (independent) are incorporated in the analysis, the more model will be complete and accurate, but only when the conditioning (independent) variables have a major role in deciding the dependent variable (Ayalew et al. 2004).

Chau and Chan (2005) describe the logistic regression as the probability of an event occurrence over the probability of non-occurrence. It is helpful for forecasting the absence or presence of a flood based on several predictor variables. The spatial prediction of a flood in logistic regression is modeled considering dependent and independent variables (Shirzadi et al. 2012). The variables could be discrete or continuous or any mixture of both types. The presence or absence of flood is represented by a binary variable (dependent variable). The logistic function is valid for flood susceptibility analysis if the dependent variables are binary (Atkinson and Massari 1998). Here, the dependent variable falls between 0 and 1, where zero represents a 0% probability (no flood) and one exhibits a 100% probability of flood occurrence (Dai et al. 2004). Quantitatively, the correlation between the flood occurrence and its dependency on other variables is described as:

$$p = \frac{1}{1 + e^{-z}} \tag{1}$$

where $p$ is the probability of flood occurrence and ranges from 0 to 1. The value $p$ is the estimated probability of flooded areas. The logistic regression considers fitting an equation of the following outline to the data:

$$Z = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots + \beta_n X_n \tag{2}$$

where $Z$ is the linear combination of the flood casual factors $X_i (i = 1, 2, \ldots\ldots, n)$, $\beta_o$ is the intercept of the model, and $\beta_i$ are the parameters of the logistic regression model.

## Results and discussions

Before beginning the logistic regression analysis, various statistical methods are employed to guarantee that the regression assumptions are valid. This is done through a series of testings including (1) Pearson's correlation between the independent variables, the (2) multicollinearity test, and (3) heteroscedasticity test.

### Statistical validation analyses

#### Pearson's correlation

Correlation matrix has been generated using Pearson's method in order to measure the correlation between the independent variable with each other, as shown in the Table 1. Each variable correlates perfectly with itself, as an evidence by the coefficients of 1.0 at the intersection of a particular variables' row and column. A poor correlation exists between most of the variables and indicates that most of the variables are independent of each other. However, a moderate correlation appears between digital elevation model (DEM) and soil type as

**Table 1** Pearson's correlation matrix among the independent conditioning variables, and VIF multicollinearity values

| | LULC | Rainfall amount | Soil type | Flow accumulation | DEM | Topographic slope |
|---|---|---|---|---|---|---|
| LULC | 1.0 | | | | | |
| Rainfall amount | − 0.0178 | 1.0 | | | | |
| Soil type | 0.218754 | − 0.355 | 1.0 | | | |
| Flow accumulation | − 0.15394 | − 0.1961 | − 0.0137 | 1.0 | | |
| DEM | 0.115976 | − 0.1861 | 0.64573 | − 0.1938 | 1.0 | |
| Topographic slope | − 0.22588 | 0.11736 | − 0.2839 | 0.05264 | − 0.0855 | 1.0 |
| VIF | 1.005 | 2.125 | 1.138 | 1.04 | 1.308 | 1.221 |

accounting for 0.64 (see Table 1). This indicates an existence of a multicollinearity between the DEM and soil type variables, as these two variables are linearly related with one another. This correlation may result in the problem of logistic linear parameter estimation.

## Multicollinearity test

Multicollinearity test is an important step in flood susceptibility mapping. Multicollinearity means an existence of a linear relationship between some or all explanatory variables of a regression model (Gujarati 2004). The presence of a linear relationship between factors can cause a division-by-zero during regression computation. This problem can cause the calculations to be false, and the logistic parameters are incorrect and or inexact; dividing by a small number will disrupt the results.

Multicollinearity is a technique in which two or more independent variables in a logistic regression model are highly correlated. It means that one variable can be predicted linearly from the others with a significant level of accuracy. Multicollinearity does not decrease the reliability and prediction power of the model. It only affects estimations related to individual predictors (Gujarati 2004).

Different methods can be used to detect multicollinearities, such as variance inflation factors (VIF), pairwise scatter plots, and eigenvalues in a correlation matrix. In this study, VIF is used to detect multicollinearity for each flood conditioning parameter. In statistics, the VIF measures the strictness of multicollinearity in terms of an ordinary least squares regression. It is represented by an index that calculates how much the variance of an estimated regression coefficient is augmented because of multicollinearity. The VIF factor for $\beta_i$ (Eq. 2) can be calculated from the following formula:

$$\mathrm{VIF}_i = \frac{1}{1 + R_i^2} \tag{3}$$

where $R_i^2$ is the coefficient of determination of the regression equation (Eq. 3). If VIF is more than 10, then multicollinearity is high (Kutner et al. 2004). If this condition is not met (i.e., VIF greater than 10), then the general linear model is not applicable and cannot be considered for the process of information estimation (Gujarati 2004).

Table 1 shows that VIF value for all independent variables is less than (5), which means that the value of VIF for all independent variables is free of a multicollinearity issue. Therefore, all of these variables must be included and tested in the logistic regression, as each variable may have an effect of the flood.

## Heteroscedasticity test

The assumption of homoscedasticity (meaning "same variance") is central to the logistic regression model. Homoscedasticity means that the error is the same across all values of the independent variables. Heteroscedasticity exists when the size of the error term varies between the values of an independent variable. In this study, Eicker–Huber–White standard errors (also Huber–White standard errors or White standard errors) will be used to calculate the standard error of the regression (Eicker 1967; Huber 1967; White 1980; Kleiber and Zeileis 2006). Heteroscedasticity-consistent (HC) standard errors are employed to permit the fitting of a model that incorporate heteroscedastic residuals. When the HC standard error is less than 5%, the regression estimation considered a heteroscedastic problem. Table 2 shows that the standard error of regression is (0.167), which is greater than (5%); this

**Table 2** Logistic regression coefficients

| Independent variables | ($\beta$) coefficient | statistics-$t$ | Sig. $R$ |
|---|---|---|---|
| Land cover-land use (LULC) | 0.1690 | 1.030 | 0.304 |
| Rainfall amounts (mm) | 0.0372 | 2.920 | 0.004 |
| Soil type | 0.0348 | 0.773 | 0.440 |
| Flow accumulation | 0.0066 | 3.310 | 0.00 |
| DEM (meters) | − 0.0075 | − 0.681 | 0.046 |
| Topographic slope (in degrees) | 1.0483 | 1.895 | 0.020 |
| C | − 1.4054 | 0.0106 | 0.991 |
| White (standard error of regression) | | 0.167 | |
| Probability ($F$-statistic) | | 0.000 | |
| $R^2$ | | 0.91 | |

indicates that the logistic regression is free from any heteroscedastic problem (Gujarati 2004).

### Logistic regression estimation

In the logistic regression, a mathematical equation was formulated. EViews 10 software is used for the logistic regression analysis. The coefficient of the related decision variable (LULC, rainfall amount, soil type, flow accumulation, DEM, and topographic slope) is shown in Table 2 and Eq. 4. These factors were used to predict the flood in GIS to obtain the probability map. A positive sign value for a specific decision variable means that the effect of the variable increases the possibility of flood occurrence. On the other hand, a negative sign means that the presence of the decision variable decreases the probability of flood occurrence (Chauhan et al. 2010).

$$\begin{aligned} Z = {} & (0.169 \times \text{LULC}) + (0.0372 \times \text{Rainfall}) \\ & + (0.0348 \times \text{Soil type}) \\ & + (0.0066 \times \text{Flow accumulation}) \\ & + (-0.0075 \times \text{DEM}) \\ & + (1.0483 \times \text{Topographic slope}) - 1.4054 \end{aligned} \tag{4}$$

The logistic regression coefficient values are shown in Table 2. The coefficient of determination ($R^2$) for the regression model was found to be 91%. This indicates that the existence of the considered independent variables explains 91% of the flood occurrence (see Table 2). Table 2 also shows that significant level (Sig. $R$) for all independent variables was calculated from the logistic regression model. The notable influence on flood occurrence can be established with the significant factor (Sig. $R$) (Papadopoulou-Vrynioti et al. 2013). If a conditioning factor has a Sig. $R$-value less than 5%, then this conditioning factor considered statistically has a high effect on the flood. The results show that the flow accumulation, topographic slope, rainfall amount, and DEM are most influencing variables with values of (0.00), (0.002), (0.004), and (0.046), respectively. Other independent variables, such as LULC, and soil type have less impact on flood occurrence as the others.

## Validation of the flood susceptibility map

In flood susceptibility analysis, it is important to locate areas that may be impacted by future floods. Regardless of the method used for validation, it is important to validate the susceptibility maps with regard to unknown future flood (Chung and Fabbri 2003). Here, the known flood areas from past years which have not been used in the model training were used to validate the flood susceptibility map (Fig. 6). For the validation technique, receiver operating characteristics (ROC) method is used to find the accuracy of the produced flood susceptibility map using the logistic regression (Pradhan and Lee 2009; Omid Rahmati et al. 2016). ROC method calculates the success and prediction rates of the flood susceptibility map based on data on previous flood events. The success rate shows how accurately the LR model performed with the training data. The prediction rate shows how accurately the model can forecast the affected flood areas. The success rate is obtained if the AUC is calculated using the training dataset. The prediction rate is calculated when the AUC is obtained using the testing dataset. Validation was conducted by comparing the flood susceptibility map using area under curve (AUC) method of the ROC against known flood dataset. The AUC of ROC illustrates the correctness of a prediction model by finding the capacity of the technique to calculate the occurrence and non-occurrence of a flood from pre-defined historical data and areas (Pourtaghi and Pourghasemi 2014). The AUC is used to assess qualitatively the flood susceptibility prediction accuracy (Pradhan 2010). The relationship between the AUC and prediction/success accuracies of the flood susceptibility map can be described into the following categories: 0.6–0.7 (average), 0.7–0.8 (good), 0.8–0.9 (very good), and 0.9–1.0 (excellent) (Yesilnacar 2005). The

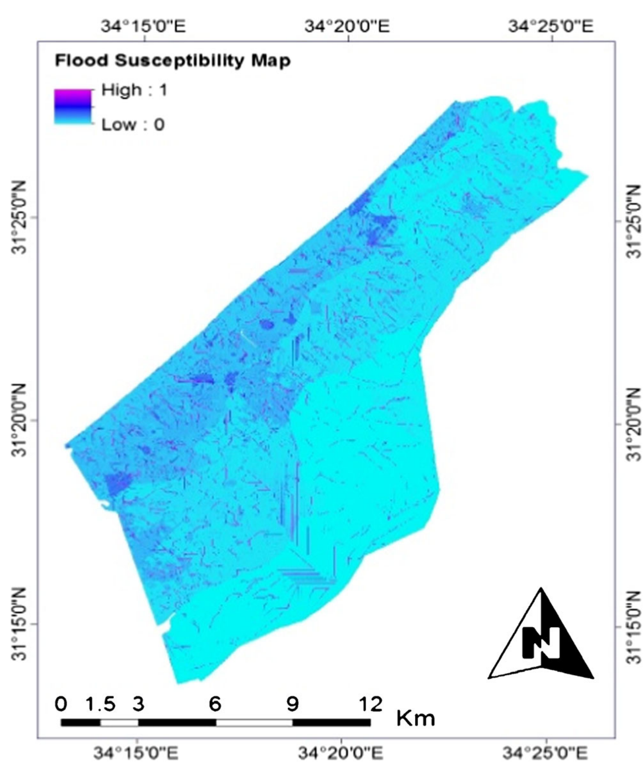



**Fig. 6** Flood susceptibility map

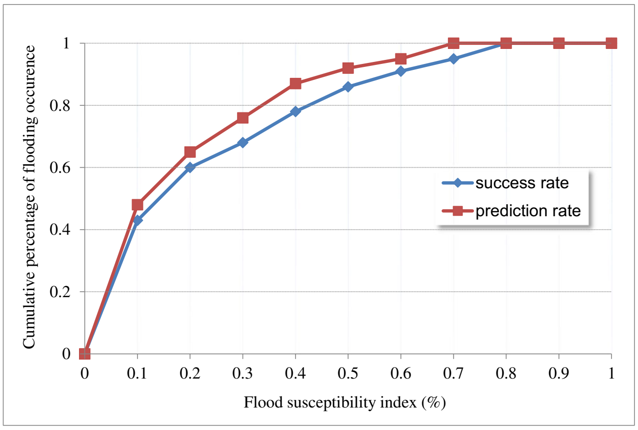



**Fig. 7** Validation of flood susceptibility map using area under curve (AUC) of receiver operating characteristics (ROC) method

ROC curve for the logistic regression is shown in Fig. 7, with AUC about 81 and 76% of prediction and success accuracies, respectively.

## Summary and concluding remarks

Flood is a major natural tragic phenomenon in the Gaza Strip, while many measures are considered to mitigate it. Therefore, predicting future flood by producing flood susceptibility mapping is an important step for reducing flooding impacts for land and watershed sustainable developments. The objectives of this paper are to: (1) produce flood susceptibility map for Gaza Strip southern watersheds using logistic regression model, (2) establish the most effective variables for the occurrence of a flood, and (3) assess the efficiency of the flood susceptibility map.

A flood inventory comprises 140 flood locations that were identified using GPS and field surveys by the Palestinian Water Authorities. The 140 locations were divided into 70% and 30% for the purpose of modeling and validation, respectively. To obtain the susceptible map, six conditioning factors (rainfall amount, soil type, land use/land cover, flow accumulation, topographic slope, and DEM) were produced using a GIS database. Three different statistical techniques were used to guarantee that the regression hypothesis is valid. The validation techniques were Pearson's correlation, multicollinearity, and heteroscedasticity tests. Logistic regression was employed to find the most influencing variables of the occurrence of a flood. Finally, the AUC method was used on the validation data to test the flood susceptibility mapping accuracy. The success and prediction rates were quantitatively measured by the logistic regression model. The largest area under curve representing the prediction and success rate were 76 and 81%, respectively. The result of the AUC indicates a very good accuracy of the flood susceptibility map.

One of the objectives of this paper is to find the most influencing variable of flood occurrence. Based on the results of the logistic regression model, the order of the flood conditioning factors was flow accumulation, topographic slope, DEM, rainfall amount, LULC, and soil type. The most significant factor is the flow accumulation, while as the least significant factors are LULC and soil type.

It was found that the southern Gaza governorates are falling under a medium susceptibility index. In general, the western part of the study area had medium–high flood susceptibility, whereas the eastern part had low–medium flood susceptibility. This paper shows that the utilization of flood susceptibility map is a useful basis in taking preventive actions to mitigate floods and land-use planning. In the areas which fall under medium–high flood susceptibility, it is necessary to avoid constructing basements or underground parking lots. Alternatively, authorities should adopt various options to keep houses safe and protect against basement flooding. However, the flood susceptibility map may be less useful on small site scale, where geographic heterogeneities may exist. It is recommended to include more validation data to improve model accuracy and generalize its applicability.

**Acknowledgments** The authors would like to thank the anonymous reviewers for their valuable comments on the manuscript.

**Authors' contributions** Ahmed E. M. Al-Juaidi conducted the validation test and prepared the manuscript. Ayman M. Nassar analyzed the GIS maps. Omar E. M. Al-Juaidi performed the statistical analysis.

## References

Adiat K, Nawawi M, Abdullah K (2012) Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool—a case of predicting potential zones of sustainable groundwater resources. J Hydrol 440–441:75–89

Aish A, De Smedt F (2004) Modeling of a groundwater mound resulting from artificial recharge in the Gaza Strip, Palestine, Proceedings of the Second Israeli-Palestinian International Conference on Water for Life in the Middle East, October 10–14, 2004, Turkey, pp 114–122

Akay H, Baduna Kocyigit M, Yanmaz AM (2018) Effect of using multiple stream gauging stations on hydrologic parameters and estimation of hydrograph of ungauged neighboring basin. Arab J Geosci 11:282

Al-Agha M (1995) Environmental contamination of groundwater in the Gaza Strip. Environ Geol 25:109–113

Al-Juaidi AE (2018) A simplified GIS based SCS-CN method for the assessment of land use change on runoff. Arab J Geosci 11:269

Al-Juaidi AE, Kaluarachchi JJ, Kim U (2010) Multi-Criteria Decision Analysis of Treated Wastewater Use for Agriculture in Water Deficit Regions. J Am Water Resour Assoc 46(2):395–411

Al-Juaidi AE, Kaluarachchi JJ, Mousa AI (2014) Hydrologic-economic model for sustainable water resources management in a coastal aquifer. J Hydrol Eng ASCE 19(11):04014020

Al-Juaidi AE, Rosenberg DE, Kalaruchchi JJ (2009) Water Management with wastewater treatment and reuse, desalination, and conveyance, to counteract climate change in the Gaza Strip. AWRA Specialty Conference on Climate Change, Anchorage, Alaska, 4–6 May 2009

Al-Juaidi AE, Rosenberg DE, Kaluarachchi JJ (2011) Water management with wastewater treatment and reuse, desalination, and conveyance to counteract future water shortages in the Gaza Strip. Int J Water Resour Environ Eng 3(12):266–282

Atkinson PM, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. Comput Geosci 24(4):373–385

Ayalew L, Yamagishi H, Ugawa N (2004) Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides 1:73–81

Baduna Kocyigit M, Akay H, Yanmaz AM (2017) Effect of watershed partitioning on hydrologic parameters and estimation of hydrograph of an ungauged basin: a case study in Gokirmak and Kocanaz, Turkey. Arab J Geosci 10:331

Chaney NW, Herman JD, Reed PM, Wood EF (2015) Flood and drought hydrologic monitoring: the role of model parameter uncertainty. Hydrol Earth Syst Sci 19:3239–3251

Chau KT, Chan JE (2005) Regional bias of landslide data in generating susceptibility maps; case of Hong Kong Island. Landslides 2:280–290

Chauhan S, Sharma M, Arora MK (2010) Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model. Landslides 7(4):411–423

Chormanski J, Okruszko T, Ignar S, Batelaan O, Rebel KT, Wassen MJ (2011) Flood mapping with remote sensing and hydrochemistry: a new method to distinguish the origin of flood water during floods. Ecol Eng 37:1334–1349

Chung CF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Nat Hazards 30:451–472

Dai FC, Lee FC, Tham LG, Ng KC, Shum WL (2004) Logistic regression modelling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong. Bull Eng Geol Environ 63:315–327

Eicker F (1967) Limit theorems for regression with unequal and dependent errors. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp 5982

Fernando N, S Grimaldi S, Santini M, Petroselli A, Ubertini L (2008) Hydrogeomorphic properties of simulated drainage patterns using digital elevation models: the flat area issue. Hydrol Sci J 53(6): 1176–1193

Fisher FM et al (2005) Liquid assets: an economic approach for water management and conflict resolution in the middle east and beyond, resources for the future. Routledge, Washington, DC, p 272

Gujarati DN (2004) Basic econometrics, 4th edn. The MacGraw Hill Company, New York City 1002 p

Helsel DR, Hirsch RM (2002) Statistical methods in water resources—hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap A3, 510 p

Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley series in probability and mathematical statistics. Wiley, New York, p 307

Hsu T, Shih D, Chen W (2014) Destructive flooding induced by broken embankments along Linbian Creek, Taiwan, during typhoon Morakot. J Hydrol Eng ASCE 20(7):05014025,1–05014025,9

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp 221–233

Iwalewa TM, Elamin AS, Kaka SI (2016) A coupled model simulation assessment of shallow water-table rise in a Saudi Arabian coastal city. J Hydro Environ Res 12:46–58

Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WN, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ Earth Sci 67:251–264

Kleiber C, Zeileis A (2006) Applied Econometrics with R. User-2006 conference. Archived from the original (PDF) on April 22, 2007

Konadu D, Fosu C (2009) Digital elevation models and GIS for watershed modelling and flood prediction—a case study of Acca Ghana. In: Appropriate Technologies for Environmental Protection in the developing world. Springer, pp 352–332

Kourgialas NN, Karatzas GP (2011) Flood management and GIS modelling method to assess flood-hazard areas—a case study. Hydol Sci J 56(2):212–225

Kutner MH, Nachtsheim CJ, Neter J (2004) Applied linear regression models, 4th edn. McGraw-Hill Irwin

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

Ohlmacher G, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansa, USA. Eng Geol 69:331–343

Omid Rahmati O, Pourghasemi HR, Zeinivand H (2016) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto Int 31(1):42–70

Palestinian Water Authority (PWA) (2012) Status report of water resources in the occupied state of Palestine, year of flooding water harvesting, Annual water resources status report, 17 p

Papadopoulou-Vrynioti K, Bathrellos GD, Skilodimou KD, Kaviris G, Makropoulos K (2013) Karst collapse susceptibility mapping considering peak ground acceleration in a rapidly growing urban area. Eng Geol 158:77–88

PEPA (1996) Gaza environmental profile, Part one-inventory of resources. Palestinian Environmental Protection Authority (PEPA), Gaza, pp 1–21

Pourtaghi ZS, Pourghasemi HR (2014) GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. Hydrogeol J 2(3):643–662

Pradhan B (2010) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J Spat Hydrol 9: 1–18

Pradhan B, Lee S (2009) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environ Earth Sci, on-line first 60: 1037–1054

Sarhadi A, Soltan S, Modarres R (2012) Probabalistic flood inundation mapping of ungauged rivers: linking GIS techniques and frequency analysis. J Hydrol 458:68–86

Schumann GJ, Bates PD, Neal JC, Andreadis KM (2014) Technology: fight floods on a global scale. Nature 507:169

Shirzadi A, Saro L, Joo OH, Chapi K (2012) A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. Nat Hazards 64:1639–1656

Tehrany MS, Lee MJ, Pradhan B, Jebur M, Lee S (2014a) Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environ Earth Sci 72:4001–4015

Tehrany MS, Pradhan B, Jebur MN (2014b) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J Hydrol 512:332–343

Tehrany MS, Pradhan B, Mansor SH, Ahmed N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125:91–101

Wagener T, Sivapalan M, Troch P, Woods R (2007) Catchment classification and hydrologic similarity. Geogr Compass 1(4):901–931

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4): 817–838

Yesilnacar EK (2005) The application of computational intelligence to landslide susceptibility mapping in Turkey [PhD thesis]. Melbourne: Department of Geomatics the University of Melbourne, 423 p

Youssef AM, Pradhan B, Sefry SA (2016) Flash flood susceptibility assessment in Jeddah city (Kindom of Saudi Arabia) using bivariate and multivariate statistical models. Environ Earth Sci 75(12)