Journal of Hydrology 504 (2013) 69–79

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS

Mahyat Shafapour Tehrany, Biswajeet Pradhan *, Mustafa Neamah Jebur

Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 11 August 2013
Received in revised form 21 September 2013
Accepted 22 September 2013
Available online 29 September 2013
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Flood susceptibility
Decision tree (DT)
Ensemble
GIS
Remote sensing
Malaysia

SUMMARY

Flood is one of the natural hazards which occur all over the world and it is critical to be controlled through proper management. Severe flood events in Kelantan, Malaysia cause damage to both life and property every year, and therefore the development of flood model to recognize the susceptible areas in watersheds is important for decision makers. Remote sensing (RS) and geographic information system (GIS) techniques could be useful in hydrological studies while they are able to fulfill all the requirements for comprehensive, rapid and accurate analysis. The aim of the current research is to compare the prediction performances of two different approaches such as rule-based decision tree (DT) and combination of frequency ratio (FR) and logistic regression (LR) statistical methods for flood susceptibility mapping at Kelantan, Malaysia. DT is based on the rules which are created precisely and strongly by considering all the characteristics of the variables which can enhance the performance of the flood susceptibility mapping. On the other hand, LR as multivariate statistical analysis (MSA) has some weak points. For that reason, FR was used to analyze the impact of classes of each variable on flood occurrence and overcome the weakness of LR. At first, flood inventory map with a total of 155 flood locations was extracted from various sources over the part of the Kelantan. Then the flood inventory data was randomly divided into a testing dataset 70% (115 flood locations) for training the models and the remaining 30% (40 flood locations) was used for validation purpose. The spatial database includes digital elevation model (DEM), curvature, geology, river, stream power index (SPI), rainfall, land use/cover (LULC), soil type, topographic wetness index (TWI) and slope. For validation both success and prediction rate curves were performed. The validation results showed that, area under the curve for the results of DT and integrated method of FR and LR was 87% and 90% for success rate and 82% and 83% for prediction rate respectively.

1. Introduction

Flood is considered as a severe natural hazard and the coverage of its damages is not measurable (Rozalis et al., 2010). Kron (2002) describes flooding as a result of heavy precipitation and snow melting that makes the rivers overflow form their normal border and temporarily covers the land which was not used to be covered by water. This type of flooding is classified as river flood. While there are two other types of flash flood and coastal flood exist, but river flood can be predicated through proper methods (Jonkman, 2005). In natural hazard management especially in flood management time is one of the most important factors i.e. the employed model should be fast in order to assist the early warning and prevention measures. Many studies have been done in order to measure and classify the flood impacts from various perspectives. Generally, damages can be direct and indirect, or tangible and intangible which all should be considered in flood damage assessment (Merz et al., 2004; Smith and Ward, 1998). Opolot (2013) stated that between 2000 and 2008 almost 99 million people per year were affected by flood alone worldwide. The frequent increase of flood events are mainly due to rapid urbanization and civilization along the rivers, and also cutting the forests (Bronstert, 2003; Christensen and Christensen, 2003). For that reason, susceptible areas to the flood should be detected in order to avoid generating more development in these areas and also to be able to have fast emergency response in various circumstances.

High frequency of the flood occurrence in Malaysia made this disaster as the most important natural hazard causing many deaths, loss of properties and damages to the ecosystem (Pradhan and Youssef, 2011). Since the 1920s many reports have been recorded about the flood occurrences in Malaysia. Department of Irrigation and Drainage (DID) stated that 9% of land area (29,800 km^2) in Malaysia is susceptible to flood and also 22% of the population

* Corresponding author. Tel.: +60 3 89466383; fax: +60 3 89466459.
E-mail addresses: biswajeet24@gmail.com, biswajeet@lycos.com (B. Pradhan).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.09.034

(4.82 million) is affected by this disaster (Pradhan, 2010a). Kelantan is one of the 13 states of Malaysia and is highly affected by the annual monsoon floods during last decades. Recently, heavy monsoons rainfall has triggered floods in Malaysia and especially threatens some states such as Kelantan, Terengganu and Pahang which are located along Malaysia's east coast (Pradhan and Youssef, 2011). The flood cost nearly million dollars of property and many lives which could have been prevented or mitigated if an early warning system was in place. Human activities such as interference in natural cycle by land use/cover (LULC) changes, unplanned urban expansion near to the bank of the rivers, and uncontrolled construction of buildings can influence the spatial and temporal pattern of hazards. Therefore, an assessment of the basin structure, climate condition, and susceptible areas, assists to prevent the damages which threat the human lives and properties.

Usually, flood management can be done through four stages: prediction, preparation, prevention and damage assessment (Konadu and Fosu, 2009). The efficiency of RS and GIS made the revolution in hydrology and specially flood management which could fulfill all the requirements of each stage. Different types of analyses can be done prior to the flood occurrence, during and after its event. Traditional flood models are increasingly improved or replaced by rule-based and automated methods which are more robust in hazard analyses (Hostache et al., 2013). So floods can be predicted and the flood risk and vulnerable areas can be mapped out. Through susceptibility analysis the areas which have high potential to the flooding can be recognized and therefore; early warning and emergency response can be performed in order to facilitate early preparations and decrease the effects of the disaster (Kia et al., 2012).

2. Previous studies

In the recent years, many methods have been developed and applied in flood susceptible mapping. Some hydrological models such as WetSpa (Liu and De Smedt, 2004), HYDROTEL (Fortin et al., 2001) and SWAT (Jayakrishnan et al., 2005) are integrated with RS and GIS for the purpose of data collection and spatial analysis. But more robust and automated methods are needed to be used in order to solve the disadvantages of the traditional hydrological methods (Li et al., 2012). Townsend and Walsh (1998) is one of the pioneer studies that proved the possibility of flood prediction through RS modeling in GIS environment. In the literature, many methods have been reported using these techniques such as Youssef et al. (2011), Pradhan (2010a), Pradhan et al. (2009), García-Pintado et al. (2013), Stephens et al. (2012), Prakash et al. (2012), Degiorgis et al. (2012), Hostache et al. (2010) and many more. Although some of them were able to produce acceptable results, still they contain some weak points that need to be improved (Matgen et al., 2007).

For instance artificial neural network (ANN) as a popular method in flood and other disaster modeling is considered as a black box which has complex procedure, needs high computer capacity and its performance is not easy to understand (Maier and Dandy, 2000; Pradhan and Buchroithner, 2010). Flood susceptibility mapping using ANN methods have been applied in various parts of the world (Dixon, 2005; Islam et al., 2001; Kia et al., 2012). Kia et al. (2012) used ANN model and simulated the flood-prone areas at Johor River Basin, Malaysia. They stated that this technique is able to deal with uncertainties in the input dataset and can extract information from incomplete or contradictory dataset (Oh and Pradhan, 2011; Pradhan, 2010a). But poor predictions can be expected when the validation data contain values outside of the range of those used for training. In the case where large number of variables are used, it makes the entire modeling process time consuming (Ghalkhani et al., 2012). However, fuzzy logic model has more transparent structure than ANN and has been employed in a variety of hydrological applications (Tilmant et al., 2002). Also some qualitative methods like analytic hierarchy process (AHP) require expert knowledge and contains many biases (Matori, 2012). This method is recommended to be used for regional studies, while the global solution and transferable method is required in flood applications (Ayalew and Yamagishi, 2005). Fernández and Lutz (2010) used multicriteria decision analysis of AHP for the purpose of flood mapping over the cities of Yerba Buena and Tucuman in Tucumán Province, Argentina. They also stated that the method can be affected by the judgment of the experts (Chang et al., 2008).

Traditional methods are mostly based on linear assumption which is not proper for catchment studies which has nonlinear structure (Liu and De Smedt, 2004). While the aforementioned methods contained some weak points, GIS analysts has been trying to develop more sophisticated and ensemble methods to minimize the disadvantages of previous methods. For instance adaptive-network-based fuzzy inference system (ANFIS) which is a combination of ANN and fuzzy interface system (FIS) is found to be optimal (Tien Bui et al., 2012b; Akgun et al., 2012). It needs minimum input from expert and performs fast, thus appears better suited to the flood environment studies. The problem is, it entails a large number of parameters which limits its use and reduce the demand due to the difficulty of data collection (Chau et al., 2005). Similarly, genetic algorithm-based artificial neural network (ANN-GA) which is a hybrid integration of ANN and genetic algorithm (GA) algorithms is another example of such modeling that takes advantage of the characteristics of both schemes (Wan et al., 2012). Although this method can increase solution stability and improve the performance of an ANN model, it requires longer computation time and additional modeling parameters are needed (Dawson et al., 2006).

For statistical approaches it is the rule to define strict assumptions prior to study which is considered as a drawbacks of such approaches (Benediktsson et al., 1990). Also, it is difficult to utilize them for real-life applications. However, statistical based logistic regression model (LR) could overcome these drawbacks and produce an easy way of analysis which does not require prior assumption and can be ensemble with other bivariate statistical analysis (BSA) methods such as frequency ratio (FR) (Ayalew and Yamagishi, 2005). Although multivariate LR is a robust method among other statistical methods, it still has some disadvantages to analyze the classes of each flood influencing parameter. To eliminate those weaknesses some studies used LR as bivariate to solve this problem, but LR has some limitations to perform BSA as it uses the classes as an indicator and does not consider it in the analysis (Süzen and Doyuran, 2004; Pourghasemi et al., 2012b, 2013a,b). Another statistical method is FR which is able to perform BSA and analyze the impact of classes of each variable on flooding. The weak point of this method is the relationship between the variables which is mostly neglected. A recent study by Lee et al. (2012) is a proof of this statement as they applied individual FR model to map the flood sustainable areas in Busan, Korea. So combination of FR and LR can produce a comprehensive method, which can analyze the influence of classes of each conditioning factor on flooding. So the current research aims to use FR to perform BSA in order to produce the information of classes of each variable, and use them as an input of LR to overcome the disadvantages of both the methods.

On the contrary, DT as an advanced method which is very new in the area of flood analysis compared to other methods such as ANN (Saito et al., 2009). DT which is based on machine learning theory and is a type of statistical analysis, unlike other statistical methods, makes no statistical assumptions and can handle data from various measurement scales. It performs fast and permits to identify homogeneous groups with various susceptibility levels. Also it can facilitate the construction of the rules for making predictions about individual cases and for complex relationships (Tien Bui et al., 2012a). Saito et al. (2007) and Pradhan (2012) are good examples of studies that utilized DT to map the susceptible areas for landslides and they indicated that DT models are efficient for this purpose. However, DT has not been used for flood susceptibility assessment.

The current research aims to perform flood susceptibility mapping for the part of the Kelantan River basin using an advanced rule-based DT and ensemble statistical method of FR and LR. The proposed methods have not applied before in flood susceptibility mapping. Therefore, it is expected that using rule based method the weak points of previous methods as reported in the literature will be solved. Also, as the LR and FR are able to generate the model separately, they were used individually in most of the studies. However, both LR and FR have some limitations that the authors believe that the results can be enhanced through their combination.

3. Study area and data

This study focuses on the part of Kelantan River Basin in North East part of Peninsular Malaysia, covering 923 km^2. Kelantan state is one of the 13 states of Malaysia and Kelantan River is the major river in Kelantan state. About 85% of the Kelantan state's has been covered by the basin.

Due to the geographical location of the study area, it attracts much tourism. Unfortunately monsoon floods affect the area every year. In this study, the flood event of November 2005 was used as flood inventory. This event was recorded as the worst flooding in Kota Bahru. It resulted in strong flooding of urban and rural areas. Based on the location and characteristic of this state and the predicted climate condition in this area, it is obvious that Kelantan is highly susceptible to the flood occurrence. Kelantan state generally contains flat slope and moderately sloping areas, but the upper catchment which is located in the southern part of this stat can be considered as major run-off zone to the Kelantan River (Pradhan, 2010a). The study area contains various soil types and fine sandy loam is the dominant soil type in major part of the catchment (Pradhan and Youssef, 2011).

3.1. Flood inventory map

The precision of the historical flood events has very high impact on the accuracy of the flood susceptibility map (Merz et al., 2007). Flooded locations were identified and an inventory map was prepared using November 2005 flood event. The inventory map was split to 70% and 30% for the purpose of training and testing respectively (Ohlmacher and Davis, 2003). The training flood locations (115 points) were selected randomly to produce the dependent data which made up by 0 and 1 values. As the value of 1 represents the existence and the value of 0 illustrates the absence of flooding over the area. Additionally, same numbers of points (115) were selected as non-flooded areas to assign the value of 0 to them. It is expected to enhance the precision of the results through the consideration of non-flooded areas in the analysis. The rest of the flood events (40 points) were used for the purpose of testing. Fig. 1 shows the map of the Kelantan area and the flooded areas.

3.2. Flood conditioning factors

The flood conditioning factors and their relationship with the flooding need to be assessed (Liu and De Smedt, 2005). In order to perform susceptibility analysis the spatial database that

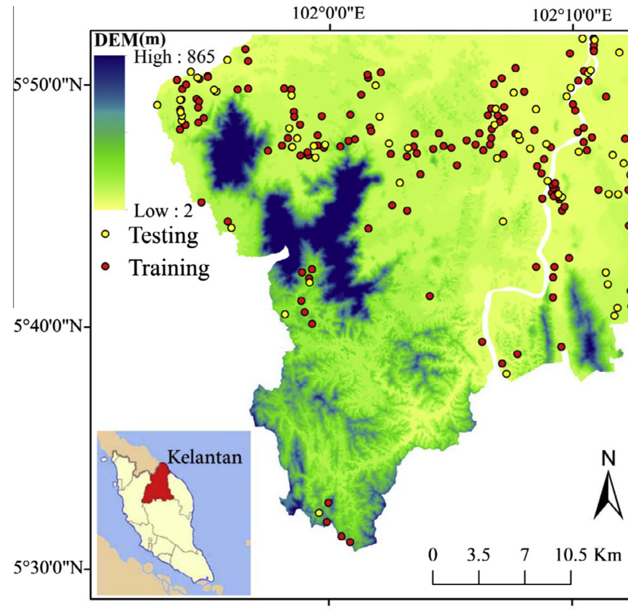

Fig. 1. Flood location map with hill-shaded map of Kelantan state, Malaysia.

contains flood conditioning factors was prepared and constructed. The selection of the conditioning factors varies form one study area to another based on different characteristics of each place. As one variable can have high degree of impact in flooding in a specific area, it can be without any influence in another region (Kia et al., 2012). For current research these variables were selected based on field survey and the information obtained from the literature (Kia et al., 2012). A total of ten conditioning factors were used in flood susceptibility mapping: digital elevation model (DEM), curvature, geology, river, stream power index (SPI), rainfall, land use/cover (LULC), soil type, topographic wetness index (TWI) and slope. Table 1 shows all the conditioning factors and their characteristics. Each factor was transformed into a grid spatial database by 15×15 m size and the grid of the Kelantan area was constructed by 2449 columns and 2622 rows (4,102,724 pixels; 923 km^2). All conditioning factors were classified using quantile method for FR modeling (Ayalew and Yamagishi, 2005).

These days, topographic data with high precision are becoming increasingly available for flood application studies in most of the countries. It has direct impact on the output of modeling and many studies limits by lack of the proper topographic data (Bates et al., 2003). This conditioning factor and its derivatives play major role to recognize the susceptible areas to flood occurrence (Pradhan, 2009). Mostly the flood prone can be detected in low elevation and flat areas. A digital elevation model (DEM) was used as a perfect source to derive topographic factors of slope, curvature, SPI and TWI. Rapid flows are made by steeper slopes, but floods tend to occur on gentle slopes. In the DEM (0–865) map, ten categories were constructed as can be seen in Fig. 2a. Curvature was classified into three classes such as concave, convex and flat (Fig. 2b). In the study area, various types of geological formations can be found in different catchments. The main lithology of the study area consists of igneous rock, limestone and sedimentary rock (Fig. 2c) (Pradhan and Youssef, 2011). Sedimentary rocks cover more than 72% of the study area making it high potential for flooding. In the case of the river map, eleven buffer categories (5 m, 15 m, 25 m, 50 m, 100 m, 200 m, 400 m, 800 m, 1000 m, 2000 m and 4000 m) were compiled using the multi-ring buffer tool in ArcGIS 10 software

Table 1
Results of ensemble FR and LR in each case.

Factor	Class	FR	Logistic coefficient
DEM	2–19	2.2	0.009
	19.1–20	1.1	
	20.1–34	2	
	34.1–39	0.9	
	39.1–40	1.5	
	40.1–60	1	
	60.1–98	0.3	
	98.1–144	0.5	
	144.1–230	0.2	
	230.1–865	0.1	
Curvature	Concave	0.8	−0.017
	Flat	1.2	
	Convex	0.8	
Geology	Igneous rocks	1.4	0.009
	Limestone	0	
	Sedimentary rocks	0.9	
River	5 m	0	0.006
	15 m	2.1	
	25 m	0.7	
	50 m	1.4	
	100 m	1.3	
	200 m	1.7	
	400 m	0.7	
	800 m	1.1	
	1000 m	0.7	
	2000 m	0.6	
	4000 m	0	
SPI	−13.8 to −10.2	1.3	0.02
	−10.3 to −7.7	0.9	
	−7.8 to −5.6	1.1	
	−5.7 to −4.3	1.6	
	−4.4 to −2.8	1.9	
	−2.9 to −1.4	0.9	
	−1.5 to −0.6	0.7	
	−0.7 to 0.1	0.3	
	0.2–1.1	0.7	
	1.2–10.8	0.4	
Rainfall	2305–2669	1.5	
	2670–2842	1.9	
	2843–2985	1.1	
	2986–3094	0.7	
	3095–3221	0.9	
	3222–3333	0.2	
	3334–3456	0.4	
	3457–3579	0.6	
	3580–3712	1.8	
	3713–3847	0.7	
LULC	Mixed horticulture	3.9	0.08
	Bush(shrubs)	0	
	Paddy	3.4	
	Mangrove	1.4	
	Urban	5.5	
	Grassland	3.7	
	Rubber	0.9	
	Cleared land	6.1	
	Forest	0.2	
	Oil Palm	0.3	
Soil types	1	3.7	0.015
	2	3.2	
	3	1.7	
	4	2.5	
	5	0	
	6	0.3	
	7	1.6	
	8	2.1	
	9	1	
	10	0.9	
	11	0.1	
	12	0	
TWI	−9.3 to −4.2	0.7	0.011
	−4.3 to −0.1	1.3	
	−0.2 to 1.8	0.8	
	1.9–2.8	0.5	
	2.9–3.9	0.5	
	4–5.4	0.8	
	5.5–7	1	
	7.1–8.4	1.3	
	8.5–9.5	1.6	
	9.6–17.2	1.7	
Slope	0	1.5	0.014
	0.1–1.5	1.2	
	1.6–2.1	1.9	
	2.2–4.3	1.2	
	4.4–8.2	0.6	
	8.3–13.2	0.5	
	13.3–17.7	0.2	
	17.8–22.1	0	
	22.2–27.3	0.5	
	27.4–87.6	0.2	

(Fig. 2d). In the SPI (−13.8 to 10.8), map, ten categories were constructed (Fig. 2e).

Rainfall data was acquired from the water level station and was categorized into ten classes (Fig. 2f). It is obvious that there is a negative relationship exists between occurrence of flooding and vegetation density. Rainfalls on the bare lands flow rapidly compared to the farmlands and forest areas. This indicates that urban areas with impervious surfaces yield more storm runoff compared to similar areas covered by mass vegetation and forest. For that reason, landuse/landcover (LULC) map was extracted from SPOT 5 imagery acquired in 2005 and classified into ten classes: mixed horticulture, brush (shrubs), paddy, mangrove, urban, grassland, rubber, clear land, forest and oil palm were constructed for this conditioning factor (Fig. 2g). Rubber covers 56% of the study area which is considered as the dominant land cover. The study area is characterized by twelve different types of soil series as can be seen in Fig. 2h)). In the TWI (−9.3 to 17.2) and slope (0–87.6) maps, ten categories were constructed for each factor (Fig. 2i, and Fig. 2j respectively). Nearly 70% of the Kelantan basin has slope angles ranging from 0° to 5° which indicates high potential for river flooding.

4. Methodology

4.1. Statistical analysis

The use of flood susceptibility maps for the purpose of LULC planning has increased significantly during the last few decades (Cerra and Prange, 2012). Such mapping assists to recognize and categorize the areas which are threatened by present or future flooding. In this paper, statistical analysis was chosen. For modeling purpose, both DT and ensemble FR and LR methods were used to compare and evaluate their efficiency in flood susceptibility mapping.

4.1.1. Bivariate statistical analysis (BSA)

FR as an efficient BSA method is used to evaluate the influence of classes of each conditioning factor on flood occurrence (Lee et al., 2012). The FR is defined as the ratio of the probability of an occurrence to the probability of a non-occurrence for given attributes (Lee and Pradhan, 2007; Pradhan et al., 2010, 2011). Usually FR is an easy method and its performance is understandable for the users (Yilmaz, 2007). The greater the FR is, the stronger

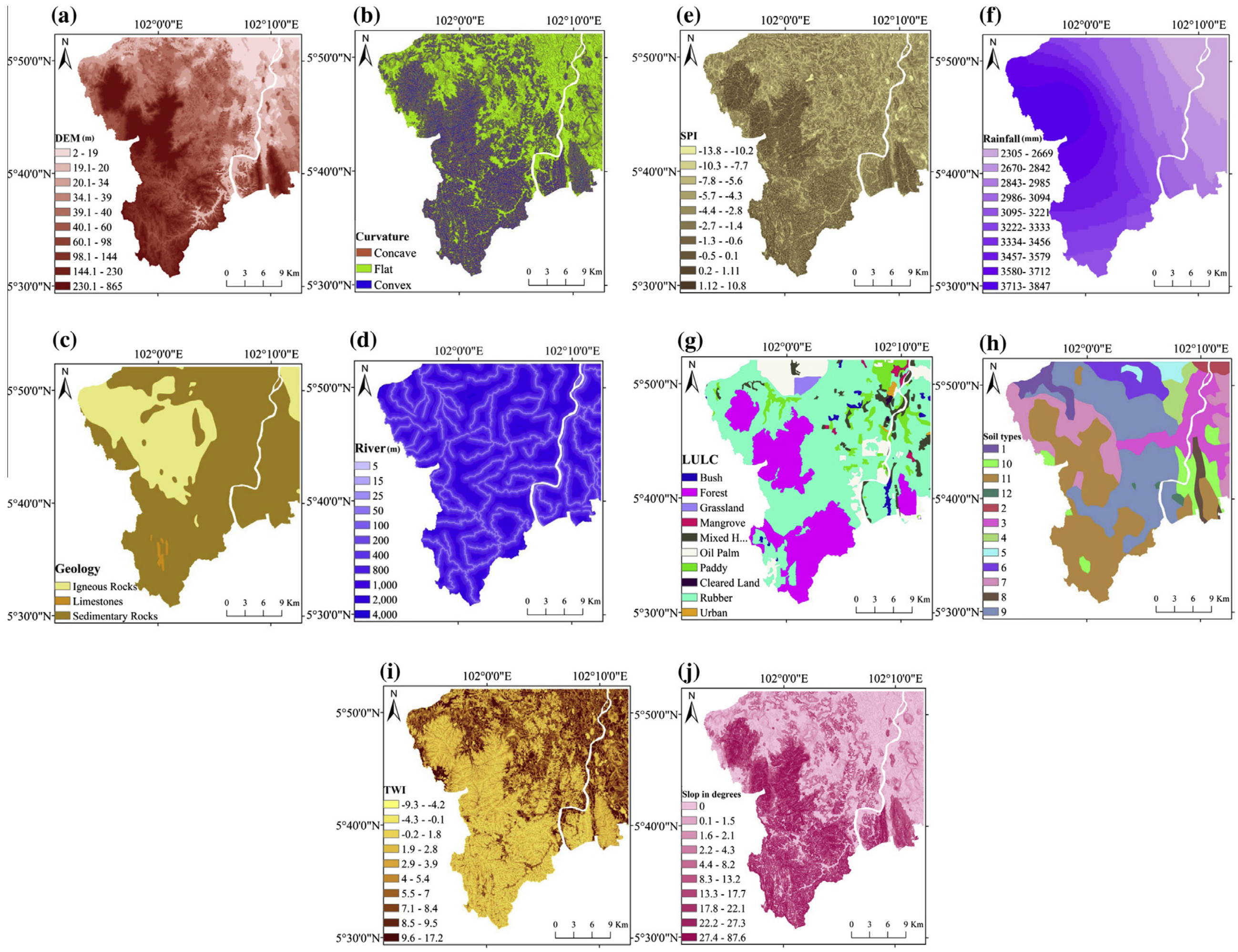

Fig. 2. Input thematic layers: (a) DEM (b) Curvature (c) Geology (d) River (e) SPI (f) Rainfall (g) LULC (h) Soil (i) TWI (j) Slope.

the relationship between flood happening and the variable (Pradhan, 2010b).

Flood conditioning factors were classified using quantile method in order to perform FR analysis in GIS (Fig. 2). FR was applied and the weights were assigned to each class of each conditioning factor (Table 1). The higher number of ratio indicates the stronger relationship between the conditioning factor and flooding and vice versa. If the FR value is higher than 1 it represents that there is strong relationship exist, and if it is less than 1 the relationship is weak (Pradhan et al., 2011). In order to use these results as an input to LR, all the acquired weights were normalized and all conditioning factors were reclassified using these normalized values.

4.1.2. Multivariate statistical analysis (MSA)

LR is a popular type of MSA which considers several parameters that may affect the probability of flood occurrence. The advantage of this method is that the data does not need to have normal distribution and the conditioning factors can be either continuous or discrete, or any combination of both types (Lee and Pradhan, 2007). The LR was first introduced by McFadden (1974) which measures the probability of any disaster in an area using specific formula that generates using the conditioning factors. This technique is capable to analyze the relationship between binary dependent variable with the scalar and nominal values as the conditioning factors (Shirzadi et al., 2012). In the current research, flood is used as dependent variable (binary) representing the presence or absence of flooding by values of 0 and 1. The LR produces the weights for each conditioning factor which can be used in GIS to produce the probability map of flood occurrence.

SPSS V.19 software was used to perform the multivariate LR analysis and coefficients were measured and listed in Table 1. The higher the logistic coefficient, the more impacts on occurrence of flooding will be expected (Ayalew and Yamagishi, 2005). The probability map can be generated using the derived logistic coefficient through the formula that can be expressed as:

$$P = 1/(1 + e^{-z}) \tag{1}$$

where p is the probability of flooding which is acquired between 0 and 1 on an S-shaped curve. Z is linear combination and it follows that LR involves fitting an equation of the following form to the data:

$$z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \tag{2}$$

where b_0 is the intercept of the model, $b_i(i = 0, 1, 2, \ldots, n)$ represents the coefficients of the LR model, and $x_i(i = 0, 1, 2, \ldots, n)$ denotes the conditioning factors (Lee and Pradhan, 2007).

4.1.3. Rule-based decision tree (DT)

DT is a powerful rule-based technique and is widely used in data classification, predictive modeling and many more (Bhaduri et al., 2008; Murthy, 1998). DT with hierarchical structure separate and classify the conditioning factors into the homogeneous groups with various susceptibility levels. DT describes the structural patterns of the dataset and their relationships as tree structures

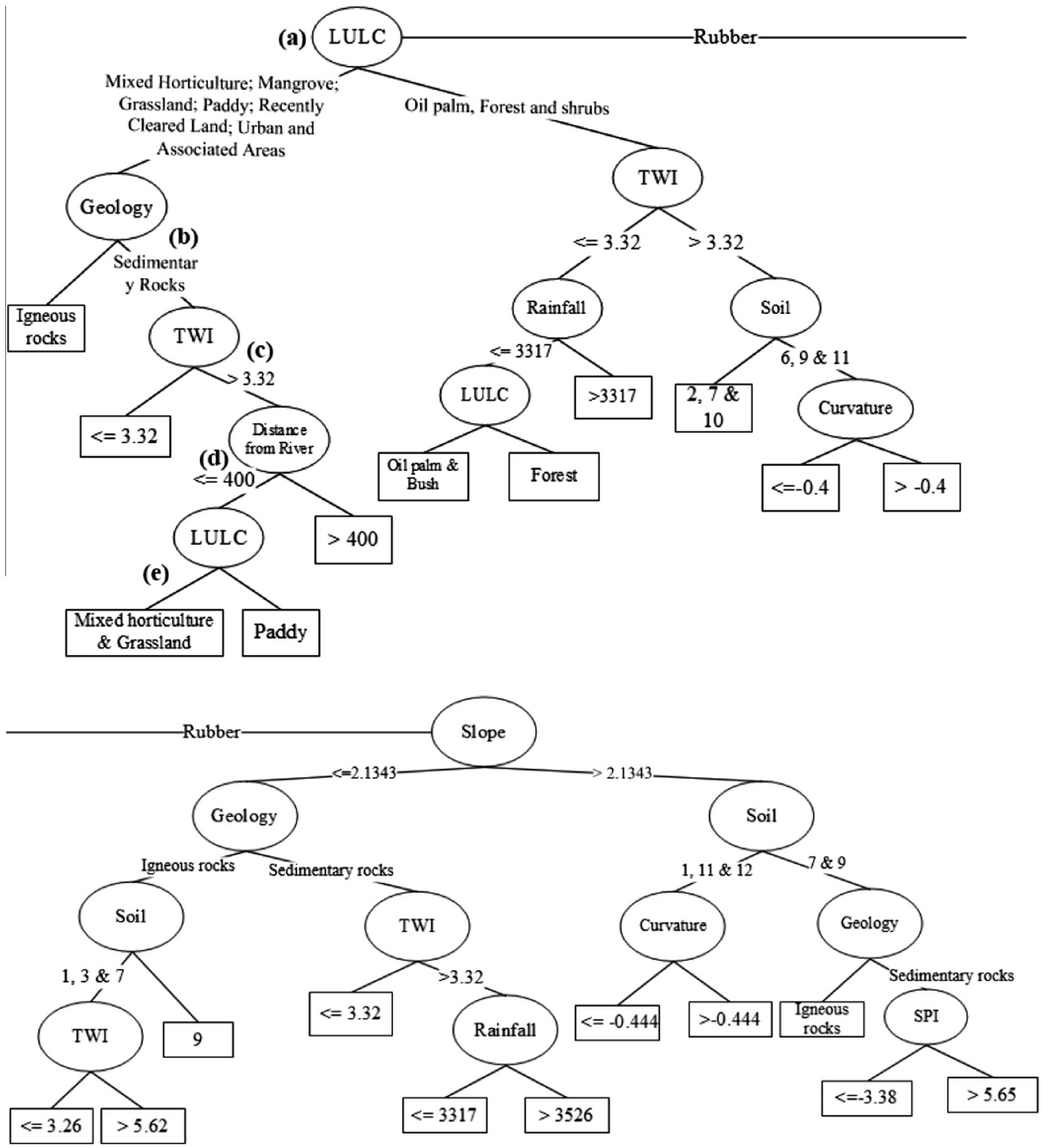

Fig. 3. DT for classifying flood susceptibility.

(Witten and Frank, 2005). Decision rules will be generated through the precise analysis in order to predict outcome from a set of input variables (Myles et al., 2004). This method is capable to model relationship between variables without having any strict assumptions about the distribution of data. Also no specific rules are needed for the data format as it can be either nominal or scalar (Bou Kheir et al., 2010). Moreover, its procedure is much easier to be understood compared to other methods like ANN (Saito et al., 2009). DT tries to achieve brief and clear representation of the relationship between dependent variable and conditioning factors (Xu et al., 2005). The tree is constructed by a root node, a set of internal nodes, and a set of terminal nodes. Each node of the tree produces binary decision that separates the classes. This analysis continues and the tree moves down until the terminal node (Schneevoigt et al., 2008). The conditioning factors that have significant impact on flooding will be used in the processing, while the others will be rejected by the program (Pal and Mather, 2003).

There are different ways to perform the DT modeling such as Chi-squared Automatic Interaction Detection (CHAID), Exhaustive CHAID, Classification and Regression Trees (CRT), Quick, Unbiased, and Efficient Statistic Tree (QUEST). (Roe et al., 2005). In this paper, CHAID was used and at each step where the conditioning (predictor) factor has the strongest relationship with the dependent variables are chosen. The classes of each conditioning factor will be merged if they are not significantly different between them with respect to the dependent variable (Berry and Linoff, 1997). Exhaustive CHAID is made by the modification of CHAID algorithm that examines all possible splits for each conditioning factor (Biggs et al., 1991). CRT first applies the segmentation on the dataset to dividing the homogenous parts based on their interaction with dependent variable. And the results of this process will extract the pure nodes (Kusiak et al., 2010). CHAID method is a proper choice for hazard modeling among others, as its performance is fast and it has the ability of multi-way node splitting (Kusiak et al., 2010). In literature, it is possible to see some applications of DT in natural hazard modeling such as landslides (Pradhan, 2013; Tien Bui et al., 2012a).

The CHAID algorithm was implemented in SPSS V.19 and the criteria were selected based on the literature (Lee and Park, 2013; Ture et al., 2009). Splitting and merging categories can be selected between the ranges of 0–1. The value of 0.9 and 0.001 were set for splitting and merging parameters which achieved through the many times trial and error. The next criterion was selected for Chi-square statistic.

5. Results and discussion

5.1. Flood susceptibility mapping using integrated FR and LR model

The first results achieved using FR method represents the weights for each classes of each conditioning factor. The FR for 15 factors was calculated from their relationship to the flood events, as shown in Table 1.

Through FR analysis the relationship between flood occurrence and the classes of each conditioning factor was found. As can be seen in the Table 1, the results of DEM analysis showed that the lowest class (2–19) was the most influential class and the highest class (230.1–865) was the least influential class of this conditioning factor on flooding. This can be proved by the natural characteristics of the floods that they mostly happen in the vast areas with low elevation, while in the peak of mountains it is not possible to be occurred. In the case of curvature, flood mostly occurred in the area that has flat curvature. In the case of geology, among the three types of geology, the ratio was highest for igneous rocks (2.97) followed by sedimentary rocks (0.9) and limestone (0). Distance from the river is one of the most important factors in the flood susceptibility mapping. Based on the results, distance of 15 m from the river had the most significant impact in flooding (2.1 ratio). Also the range of 50–100 m showed the ratio more than one which represented the long coverage and spread alluvial flood over the Kelantan area. Also the impact of the SPI was evaluated in flood occurrence. The ratio was highest (1.9) when SPI was −4.4 to −2.8, and was lowest (0.3) when SPI was in the range −0.7 to 0.1. The highest ratio for rainfall classes as a main conditioning factor of flooding was related to the area by 2670–2842 mm precipitation by the ratio of 1.9. Also the range of 3580–3712 for rainfall, showed the ratio more than 1.8.

There is a strong relationship exist between LULC and flooding. The analysis produced the highest FR ratio of 6.1 for clear area and the lowest ratio of 0 for the areas which covered by shrubs. In bare lands as there is no vegetation coverage to control and prevent the rapid flow of water on the ground, this condition increases the potential of flooding over the area as has been proved by the current results. While the areas which covered by shrubs were able to mitigate and stop flooding. Urban areas are mostly covered by the impervious surfaces such as asphalt which is strongly susceptible to the flooding. FR analysis produced the ratio of 5.5 for urban areas which shows a strong relationship with flooding. For the soil conditioning factor, the highest ratio was achieved for the soil types of 1 and 2 with ratios 3.7 and 3.2 respectively. Classes of 5 and 12 of soil types showed 0 which represents there is no important relationship exists among them and flood occurrence. In the case of TWI, the FR ratio was highest (1.7) for the class of 9.6–17.2, and it was lowest (0) for the class of >17.3. Slope results can be used as another indicator that proved the existence of strong relationship between flooding with low elevation and flat areas. FR ratio more than 1 illustrated that flood mostly happened in the areas with the slope range of 0–4.3°. Also the ratio was less than 1 and near to zero for the steeper areas with 4.4–87.6° degree of slope. It can be stated that as the slope is higher the occurrence of flooding in lower parts of the catchment will be increased as well.

The LR model was constructed based on the conditioning factors that have been reclassified using the weights achieved from FR method (Table 1). The relationship between the flood occurrence and flood influencing parameters was extracted in SPSS V.19 software. The coefficients and significances were measured using LR and are listed in Table 1. Among the conditioning factors, just curvature showed the negative relationship (−0.017) with flooding, while the rest of the conditioning factors had positive correlation with flood occurrence. The highest logistic coefficient (0.08) was assigned to the LULC as it had the most significant impact on flooding. In order to produce probability map the values obtained from LR were transferred to the ArcGIS 9.3 software and Eq (2) was applied to the conditioning factors.

$$\begin{aligned} Z = &-4.979 + (0.009\ \text{DEM}) + (-0.017\ \text{curvature}) \\ &+ (0.009\ \text{geology}) + (0.08\ \text{LULC}) + (0.006\ \text{river}) \\ &+ (0.017\ \text{slope}) + (0.015\ \text{soil}) + (0.02\ \text{SPI}) \\ &+ (0.007\ \text{rainfall}) + (0.011\ \text{TWI}) \end{aligned} \tag{3}$$

In the next step, from Eq. (1), probability index was calculated which has the range from 0 to 0.99. Fig. 4 shows the thematic map of the flood probability. This index represents the predicted probabilities of flood for each pixel in the presence of a given set of conditioning factors. Susceptibility map was prepared through the popular method adopted in the literature by dividing the probability map into specific number of classes (Ayalew and Yamagishi, 2005). The range of values of flood probability map was classified using quantile method into five categories of very low (0–0.27),

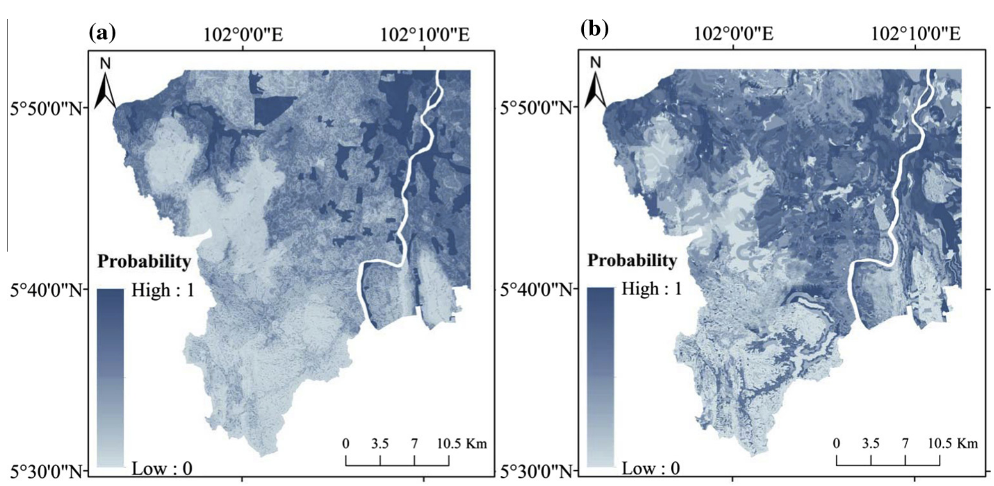

Fig. 4. The probability map obtained from: (a) ensemble FR and LR analyses and (b) DT.

low (0.28–0.5), moderate (0.51–0.76), high (0.77–0.94) and very high (0.95–0.99). The flood susceptibility map produced using the ensemble FR and LR model is shown in Fig. 5.

5.2. Flood susceptibility mapping using DT model

Using DT, rules were applied to dataset and visualized as a tree structure (Fig. 3). One of the characteristics of the robust method is using less number of parameters be able to perform modeling and generate the susceptibility map with high accuracy. As all the conditioning factors that have been entered as an input into the modeling does not have significant impact on flooding, so DT selected as the most important ones and hence generated the model based on them. In some cases two variables have same or almost similar impact on occurrence of flood; one of them is enough for analysis. So the program performs the modeling using one of them. In the current research, DEM was rejected to be used in DT modeling and the rest of the other nine conditioning factors were utilized for this purpose. In order to describe the relationships among flood susceptibility and conditioning factors, the structure of the DT was analyzed. The tree is composed of 9 variables and 22 leaves as each leaf illustrates the flood susceptibility (Fig. 3). The top-down structure of the DT represents the conditioning factors that are higher in the order of the tree structure with more significant impact on flooding than the ones which placed in the lower order.

Through this characteristic of DT, the most important conditioning factors contributed in high flood susceptibility of Kelantan catchments were found. These conditioning factors consist of all the classes of LULC except vegetation and geology of sedimentary rock, and TWI exceeding 3.32. This decision root continued by the distance from river less than 400, and mix horticulture and grass of LULC. The area that had these characteristics was found to be 100% susceptible to flooding. This area is located in the north part of the catchment (flat part of catchment). This branch can be inferred as; the first decision showing high flood susceptibility of Kelantan that tends to have all the LULC except the vegetation classes (Fig. 3a). If this condition is fulfilled, the catchment is assigned to the first left branch. Leaves under the mentioned left branch had higher flood susceptibility. The second important conditioning factor was geology with sedimentary rock (Fig. 3b). This second decision means, if the dominant geology of the catchment is sedimentary rock, the flood susceptibility will be increased. This area can be seen in the northern part of the catchment. A third decision of TWI with more than 3.32 was examined for the entire catchment with sedimentary rock (Fig. 3c). The areas that contained all three decisions up to this level, they were classified as very high susceptible to flooding. The fourth decision (distance from river less than 400, Fig. 3d) showed that catchment with frequent flood occurrence mostly happened in current areas which was located in the southern most part of Kelantan. And the last decision was the LULC of mix Horticulture and grassland (Fig. 3e).

The second important branch was related to the rubber class of LULC with slope less than 2.134, and igneous rock. This area contained the soil types of 1, 3 and 7, and TWI less than 3.328. The area which had these characteristics was classified as 83% susceptible to flood occurrence (Fig. 3). The next important branch was related to rubber class of LULC with slope less than 2.13. Similarly sedimentary rock and TWI exceeding 3.32 and rainfall more than 3317 were other characteristics of the mentioned area. The area that contained all the mentioned decisions was categorized as 82% susceptible to flooding (Fig. 3). The flood susceptibility of 0 was achieved for the area which had LULC of oil palm, forest and brush (shrubs). This region had TWI exceeding 3.32, soil types of 6, 9 and 11, and curvature exceeding −0.4 (Fig. 3). It can be concluded that the areas covered by the vegetation and forests were classified as the least susceptible areas to flood occurrence as they can control and mitigate the flood.

Fig. 4 depicts the flood probability map derived from DT rules. The probability of the catchments was then classified into five classes by adopting the same procedures applied in FR and LR models. Fig. 5 shows the flood susceptibility map of Kelantan using DT rule-based method.

Qualitatively, all thematic maps have acceptable and representable appearance. Visually, the highest flood susceptibly of Kelantan catchment is located on the northern east and northern west part of the study area. The ensemble FR and LR method produced more fine output compared to DT. However, some differences can be seen among their results, as the upper catchment classifies as very low susceptible from ensemble FR and LR method, but DT recognized some portion of upper catchment as high susceptible to flooding. But generally, both the models have agreement about the very high susceptible areas which is located along the Kelantan River. One of the advantages of DT is that there is no classification required for conditioning factors, whereas for FR method, reclassification before and after analysis was needed. This can be considered as one of the characteristics of DT that leads to have fast performance and can make it

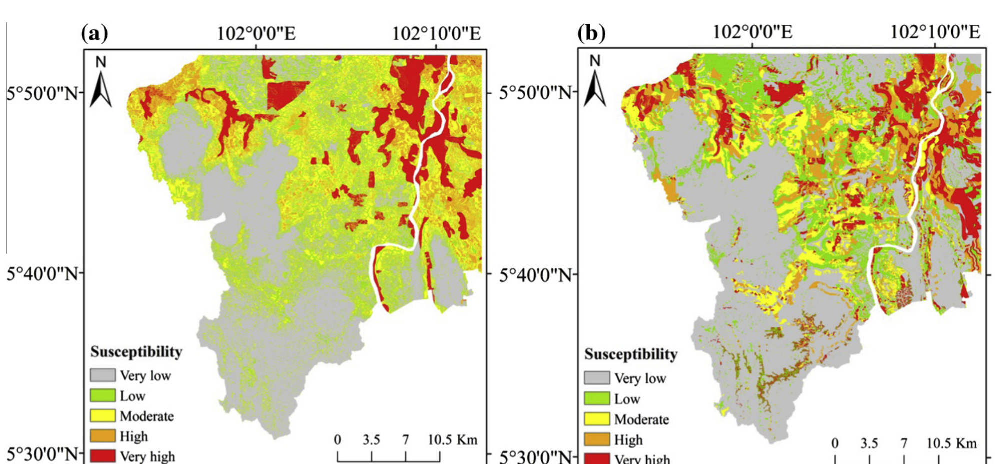

Fig. 5. Flood susceptibility map produced from: (a) ensemble FR and LR model and (b) DT.

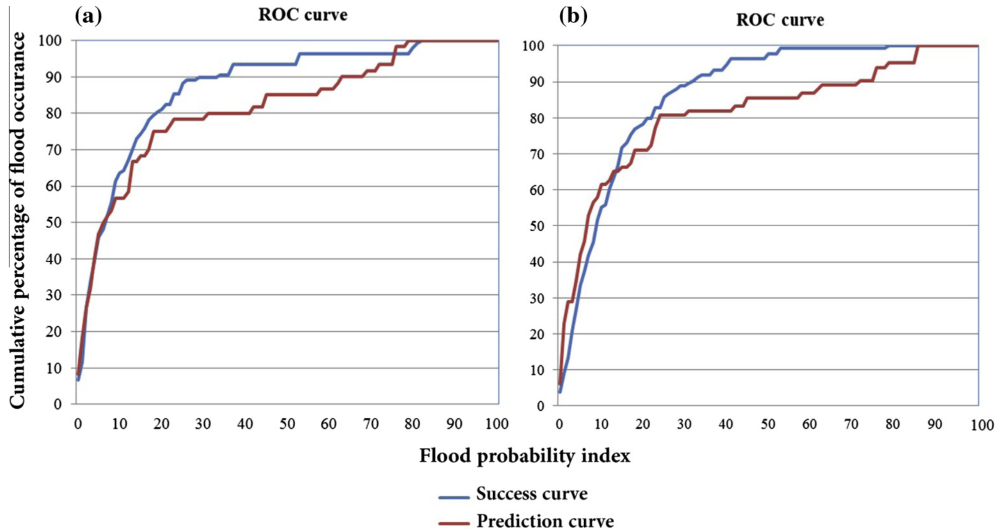

Fig. 6. ROC curve (a) the ensemble FR and LR method (b) for DT method.

optimal in natural hazard modeling. As it is not scientific to do any judgment without considering and evaluating the statistics, validation have been done using area under curve (AUC) method which is explained in the next section.

5.3. Model validation

The receiver operating characteristic (ROC) curves are very popular method due to their comprehensive, understandable and visually attractive way of accuracy assessment. Prediction and success rate should be evaluated as an essential outcome of every program (Pourghasemi et al., 2012a). The validation process was implemented by comparing the existing flood data with the acquired flood probability map. The model was validated using quantitative method of ROC curve in which the basis of the assessment is the true- and false-positive rates (Tien Bui et al., 2012b,c). The range of the AUC varies from 0.5 to 1.0, as the value of 1.0 represents the highest accuracy showing that the model was completely satisfied to predict the disaster occurrence without any biased effect (Pradhan and Lee, 2010). So the closer the value of AUC to the 1.0, the model is considered to be more precise and trustable. Also the success rate result was obtained using the training dataset i.e. 70% of the inventory flood locations (115 points). The success rate curves can be seen in Fig. 6. As the training flood layer was used to generate the model, therefore this data cannot be used for the purpose of validation, as it does not represent the real efficiency of the developed model. So prediction capability of model cannot be achieved using the training data. The prediction rate shows how well the model can predicts the flooding in an area (Tien Bui et al., 2012b). For that reason, 30% (40 points) of flooded locations that was saved for the purpose of testing were used to measure the prediction rate. So the model's generalization ability can be measured through this method. The ensemble FR and LR model produced the value of 0.906 for area under the curve (AUC) which represents 90.6% success accuracy (Fig. 6a). The prediction curve is illustrated in Fig. 6 as it shows the AUC of 0.835, which corresponds to the prediction accuracy of 83.5%. The AUC for DT method showed 0.87 and 0.82 for success and prediction accuracies, respectively (Fig. 6b).

6. Conclusion

Flood is one of the serious catastrophes that the human society is facing for many decades. Many attempts have been made in order to control and mitigate it during the last decades. Most of such actions and strategies are practiced in developed countries in which large amount of spatial database are available for analysis. Susceptible areas to the flooding should be detected to predict its spatial distribution for future events. This study addressed flood susceptibility mapping using data mining method of DT and ensemble method of FR and LR which are traditional statistical methods. The aim of this study was to assess efficiency of these two models and compare their performance to map the susceptible areas along the Kelantan River and surrounding areas in Malaysia.

For flood susceptibility mapping, a total of 10 conditioning factors were prepared and flooded inventory map was used to make the flood dichotomous dependent layer. Both models were applied and the relationship between the flood occurrence and each variable was assessed. DT rejected to utilize DEM in its performance, as other conditioning factors were existed in the dataset which had similar impact on flooding i.e. same as DEM. The ensemble method of FR and LR showed that there is negative relationship existed between the proposed conditioning factor and the flooding. Also LR revealed that LULC as the most influential conditioning factor. Generally the areas receiving higher rainfall with less vegetation and lower elevation are recognized as very high degree of susceptibility to flooding. Flood susceptibilities were then mapped for each method and AUC was used to validate the results. Both methods detected north part of catchment as the most susceptible area to flooding. The high and very high susceptibility zones derived by ensemble FR and LR made up 19.5% of the total study area.

The results from this study demonstrated that the use of an ensemble FR and LR methods could solve the weak points of each method. By using individual FR the lack of relationship between each conditioning factor with flooding reduce the efficiency of this method for MSA. And LR has some disadvantages to perform BSA. The current research could overcome these disadvantages and produced more precise method for flood susceptibility analysis by integrating both the methods. To quantitatively compare the result, the AUC was performed for both methods. The verification results showed that the area under the curve for flood susceptibility map acquired from the proposed ensemble method is 90.6% and 83.5% for success and perdition rates respectively. From the results, it is obvious that, the proposed method could enhance the results of previous works which used individual FR or LR methods.

The DT was applied and conditioning factors were ordered based on their importance. The results achieved from current research proved that DT is an efficient tool in flood modeling and susceptibility mapping. The validation process showed that DT could recognize flood prone areas by 87% and 82% success rate and prediction rate respectively. Although the accuracy of the results from DT was less that the ensemble FR and LR, it has been found that DT has so many advantages over other statistical methods, as it performs fast, makes no statistical assumptions, and is able to process data by different measurement scales. The tree structure of DT is another advantage of this method as the most significant and influential conditioning factors can be found in the initial nodes, while as it moves down in the tree the level of importance will be decreased. It means the most important conditioning factors can be found in the first levels of the tree. Another benefit of this machine learning method was its ability to select the significant conditioning factors in flooding and reject others during the model building process. In the case that two conditioning factors had same or near impact, one of them was selected to be used in analysis and the other one was left out. As a drawback of DT it can be stated that this method is susceptible to noisy data and that multiple output attributes are not allowed.

So it is recommended that the traditional hydrological methods be complemented by the new technologies of RS and GIS which can cover all the requirements of any application through their capabilities such as data collection, storage, analysis, manipulation and modeling. Similarly, it can also save time, budget and human power. The current research could detect and present the most significant conditioning factors in flooding and the most susceptible areas have been recognized. The results produced in this paper can be used to mitigate flooding in order to prevent future damages that affect the people and facilities. Also the information derived from current research can assist governments, planners, and researchers to perform proper actions in order to prevent and mitigate this disaster in future. For future studies, the authors aim to produce an automatic DT method which will be able to apply the specific rules for different areas. Also it is expected that through the combination of DT with other statistical or fuzzy methods, the efficiency of the proposed methods may increase.

References

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C., Pradhan, B., 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput. Geosci. 38, 23–34.

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65, 15–31.

Bates, P.D., Marks, K.J., Horritt, M.S., 2003. Optimal use of high-resolution topographic data in flood inundation models. Hydrol. Process. 17, 537–557.

Benediktsson, J.O.N.A., Swain, P.H., Ersoy, O.K., 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Trans. Geosci. Remote Sens. 28, 540–552.

Berry, M.J., Linoff, G., 1997. Data Mining Techniques: For Marketing, Sales, and Customer Support. John Wiley & Sons, Inc., New York, NY, USA, pp. 444.

Bhaduri, K., Wolff, R., Giannella, C., Kargupta, H., 2008. Distributed decision-tree induction in peer-to-peer systems. Stat. Anal. Data Min. 1, 85–103.

Biggs, D., De Ville, B., Suen, E., 1991. A method of choosing multiway partitions for classification and decision trees. J. Appl. Stat. 18, 49–62.

Bou Kheir, R., Greve, M.H., Abdallah, C., Dalgaard, T., 2010. Spatial soil zinc content distribution from terrain parameters: a GIS-based decision-tree model in Lebanon. Environ. Pollut. 158, 520–528.

Bronstert, A., 2003. Floods and climate change: interactions and impacts. Risk Anal. 23, 545–557.

Cerra, N., Prange, Z., 2012. Northwest arkansas comprehensive environmental analysis and regional smart growth plan. Inqiry 13, 4–22.

Chang, N.B., Parvathinathan, G., Breeden, J.B., 2008. Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region. J. Environ. Manage. 87, 139–153.

Chau, K.W., Wu, C.L., Li, Y.S., 2005. Comparison of several flood forecasting models in Yangtze River. J. Hydraul. Eng. 10, 485–491.

Christensen, J.H., Christensen, O.B., 2003. Climate modelling: severe summertime flooding in Europe. Nature 421, 805–806.

Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y., Wilby, R.L., 2006. Flood estimation at ungauged sites using artificial neural networks. J. Hydrol. 319, 391–409.

Degiorgis, M., Gnecco, G., Gorni, S., Roth, G., Sanguineti, M., Taramasso, A.C., 2012. Classifiers for the detection of flood-prone areas using remote sensed elevation data. J. Hydrol. 470–471, 302–315.

Dixon, B., 2005. Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: a GIS-based sensitivity analysis. J. Hydrol. 309, 17–38.

Fernández, D.S., Lutz, M.A., 2010. Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Eng. Geol. 111, 90–98.

Fortin, J.P., Turcotte, R., Massicotte, S., Moussa, R., Fitzback, J., Villeneuve, J.P., 2001. Distributed watershed model compatible with remote sensing and GIS data. II: Application to Chaudière watershed. J. Hydrol. Eng. 6, 100–108.

García-Pintado, J., Neal, J.C., Mason, D.C., Dance, S.L., Bates, P.D., 2013. Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling. J. Hydrol. 495, 252–266.

Ghalkhani, H., Golian, S., Saghafian, B., Farokhnia, A., Shamseldin, A., 2012. Application of surrogate artificial intelligent models for real-time flood routing. Water. Environ. J., <http://dx.doi.org/10.1111/j.1747-6593.2012.00344.x>.

Hostache, R., Lai, X., Monnier, J., Puech, C., 2010. Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River. J. Hydrol. 390, 257–268.

Hostache, R., Chini, M., Matgen, P., Giustarini, L., 2013. A new automatic SAR-based flood mapping application hosted on the European Space Agency's grid processing on demand fast access to imagery environment. In: EGUGA Conference Abstracts, Vienna, Austria. p. 13927.

Islam, M.M., Sado, K., Owe, M., Brubaker, K., Ritchie, J., Rango, A., 2001. Flood damage and management modelling using satellite remote sensing data with GIS: case study of Bangladesh. IAHS Publication, pp. 455-457.

Jayakrishnan, R., Srinivasan, R., Santhi, C., Arnold, J.G., 2005. Advances in the application of the SWAT model for water resources management. Hydrol. Process. 19, 749–762.

Jonkman, S.N., 2005. Global perspectives on loss of human life caused by floods. Nat. Hazards. 34, 151–175.

Kia, M.B., Pirasteh, S., Pradhan, B., Mahmud, A.R., Sulaiman, W.N.A., Moradi, A., 2012. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. Environ. Earth Sci. 67, 251–264.

Konadu, D.D., Fosu, C. (Eds.), 2009. Digital Elevation Models and GIS for Watershed Modelling and Flood Prediction – A Case Study of Accra Ghana, Appropriate Technologies for Environmental Protection in the Developing World. Springer, pp. 325–332.

Kron, W., 2002. Keynote lecture: Flood risk = hazard × exposure × vulnerability. In: Proceedings of the Flood Defence. Scieence Press, New York Ltd., ISBN 1-880132-54-0.

Kusiak, A., Li, M., Zhang, Z., 2010. A data-driven approach for steam load prediction in buildings. Appl. Energy 87, 925–933.

Lee, S., Park, I., 2013. Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines. J. Environ. Manage. 127, 166–176.

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

Lee, M.J., Kang, J.E., Jeon, S., 2012. Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS). IEEE, pp. 895–898.

Li, X.H., Zhang, Q., Shao, M., Li, Y.L., 2012. A comparison of parameter estimation for distributed hydrological modelling using automatic and manual methods. Adv. Mater. Res. 356, 2372–2375.

Liu, Y.B., De Smedt, F., 2004. WetSpa extension, a GIS-based hydrologic model for flood prediction and watershed management. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel. Vrije Universiteit Brussel: Brussels.

Liu, Y.B., De Smedt, F., 2005. Flood modeling for complex terrain using GIS and remote sensed information. Water Resour. Manage. 19, 605–624.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw. 15, 101–124.

Matgen, P., Schumann, G., Henry, J.B., Hoffmann, L., Pfister, L., 2007. Integration of SAR-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management. Int. J. Appl. Earth Obs. Geoinf. 9, 247–263.

Matori, A.N., 2012. Detecting Flood susceptible areas using GIS-based analytic hierarchy process. In: International Conference on Future Environment and Energy, Singapore.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

Merz, B., Kreibich, H., Thieken, A., Schmidtke, R., 2004. Estimation uncertainty of direct monetary flood damage to buildings. Nat. Hazard. Earth Sys. 4, 153–163.

Merz, B., Thieken, A.H., Gocht, M., 2007. Flood Risk Mapping at the Local Scale: Concepts and Challenges, Flood Risk Management in Europe. Springer, Netherlands, pp. 231–251.

Murthy, S.K., 1998. Automatic construction of decision trees from data: A multi-disciplinary survey. Data Min. Knowl. Disc. 2, 345–389.

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemometr. 18, 275–285.

Oh, H.J., Pradhan, B., 2011. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. Comput. Geosci. 37, 1264–1276.

Ohlmacher, G.C., Davis, J.C., 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Eng. Geol. 69, 331–343.

Opolot, E., 2013. Application of remote sensing and geographical information systems in flood management: a review. Res. J. Appl. Sci. Eng. Technol. 6, 1884–1894.

Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. 86, 554–565.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., 2012a. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat. Hazards 63, 965–996.

Pourghasemi, H.R., Mohammadi, M., Pradhan, B., 2012b. Landslide susceptibility mapping using index of entropy and conditional probability models at Safarood Basin, Iran. Catena 97, 71–84, <http://dx.doi.org/10.1016/j.catena.2012.05.005>.

Pourghasemi, H.R., Moradi, H.R., Fatemi Aghda, S.M., Gokceoglu, C., Pradhan, B., 2013a. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi criteria evaluation models (North of Tehran, Iran). Arab. J. Geosci., <http://dx.doi.org/10.1007/s12517-012-0825-x>.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., Moezzi, K.D., 2013b. A comparative assessment of prediction capabilities of Dempster-Shafer and weights of evidence models in landslide susceptibility mapping using GIS. Geomatics, Natural Hazards and Risk 4, 93–118, <http://dx.doi.org/10.1080/19475705.2012.662915>.

Pradhan, B., 2009. Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. Cent. Eur. J. Geosci. 1, 120–129.

Pradhan, B., 2010a. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. J. Spatial. Hydrol. 9.

Pradhan, B., 2010b. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J. Indian Soc. Remote Sens. 38, 301–320.

Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput. Geosci. 51, 350–365.

Pradhan, B., Buchroithner, M.F., 2010. Comparison and validation of landslide susceptibility maps using an artificial neural network model for three test areas in Malaysia. Environ. Eng Geosci. 16, 107–126.

Pradhan, B., Lee, S., 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides 7, 13–30.

Pradhan, B., Youssef, A.M., 2011. A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia. J. Flood risk. Manage. 4, 189–202.

Pradhan, B., Shafiee, M., Pirasteh, S., 2009. Maximum flood prone area mapping using RADARSAT images and GIS: Kelantan river basin. Int. J. Geoinform. 5, 11.

Pradhan, B., Youssef, A.M., Varathrajoo, R., 2010. Approaches for delineating landslide hazard areas using different training sites in an advanced artificial neural network model. Geo-Spatial Inform.Sci. 13, 93–102.

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. Int. J. Remote Sens. 32, 4075–4087.

Prakash, S., Mahesh, C., Gairola, R.M., Buyantogtokh, B., 2012. A feasibility of six-hourly rainfall forecast over central India using model output and remote sensing data. Int. J. Hydrol Sci. Technol. 2, 138–152.

Roe, B.P., Yang, H.J., Zhu, J., Liu, Y., Stancu, I., McGregor, G., 2005. Boosted decision trees as an alternative to artificial neural networks for particle identification. Nucl. Instrum. Methods A 543, 577–584.

Rozalis, S., Morin, E., Yair, Y., Price, C., 2010. Flash flood prediction using an uncalibrated hydrological model and radar rainfall data in a Mediterranean watershed under changing hydrological conditions. J. Hydrol. 394, 245–255.

Saito, H., Nakayama, D., Matsuyama, H., 2007. Estimation and evaluation of drainage basins with landslides using decision tree technique and ASTER data. J. Jpn. Landslide Soc. 44, 1–14.

Saito, H., Nakayama, D., Matsuyama, H., 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. Geomorphology 109, 108–121.

Schneevoigt, N.J., van der Linden, S., Thamm, H.P., Schrott, L., 2008. Detecting alpine landforms from remotely sensed imagery. A pilot study in the Bavarian Alps. Geomorphology 93, 104–119.

Shirzadi, A., Saro, L., Joo, O.H., 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.

Smith, K., Ward, R., 1998. Floods: Physical Processes and Human Impacts. Wiley, Chichester, p. 394.

Stephens, E.M., Bates, P.D., Freer, J.E., Mason, D.C., 2012. The impact of uncertainty in satellite data on the assessment of flood inundation models. J. Hydrol. 414, 162–173.

Süzen, M.L., Doyuran, V., 2004. A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environ. Geol. 45, 665–679.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012a. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. Math. Probl. Eng. <http://dx.doi.org/10.1155/2012/974638>.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012b. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. Comput. Geosci. 45, 199–211.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012c. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena 96, 28–40.

Tilmant, A., Vanclooster, M., Duckstein, L., Persoons, E., 2002. Comparison of fuzzy and nonfuzzy optimal reservoir operating policies. J. Water Resour. Plan. Manage. 128, 390–398.

Townsend, P.A., Walsh, S.J., 1998. Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing. Geomorphology 21, 295–312.

Ture, M., Tokatli, F., Kurt Omurlu, I., 2009. The comparisons of prognostic indexes using data mining techniques and Cox regression analysis in the breast cancer data. Expert Syst. Appl. 36, 8247–8254.

Wan, Z., Wang, G., Sun, B., 2012. A hybrid intelligent algorithm by combining particle swarm optimization with chaos searching technique for solving nonlinear bilevel. Swarm and Evolutionary Computation 8, 26–32.

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, p. 525.

Xu, M., Watanachaturaporn, P., Varshney, P.K., Arora, M.K., 2005. Decision tree regression for soft classification of remote sensing data. Remote Sens. Environ. 97, 322–336.

Yilmaz, I., 2007. GIS based susceptibility mapping of karst depression in gypsum: a case study from Sivas basin (Turkey). Eng. Geol. 90, 89–103.

Youssef, A.M., Pradhan, B., Hassan, A.M., 2011. Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. Environ. Earth Sci. 62, 611–623.