RESEARCH   **Open Access**

Geospatial approach for assessment of vulnerability to flood in local self governments

S. Deepak[*] , Gopika Rajan and P. G. Jairaj

Abstract

Recent years have shown a significant increase in the occurrence of floods globally, with an impact on habitation and different sectors of the economy. This, in turn, necessitates the use of different flood mitigation strategies, wherein flood vulnerability assessment plays a significant role. The proposed work presents a methodology that combines vulnerability under physical-environmental and socio-economic domains to assess the overall flood vulnerability at the local self-government level. The methodology was illustrated to the case of Aluva municipality, located on the banks of River Periyar, in Kerala state, India. The spatial variation of hazard inducing factors and population statistics were analysed using Geographic Information System (GIS) tools. The machine learning algorithm, Random Forest, which uses hazard inducing factors as input was implemented for the evaluation of physical-environmental vulnerability. The social vulnerability of the region was analysed using the GIS Multi-criteria decision analysis approach (MCDA), with criteria weights to incorporate the interests of different stakeholders. The critical combinations of the two domains of vulnerability in the assessment of the vulnerability to flood, to have efficient flood management in local self-government was demonstrated in this study and can be made use of for any flood event.

Keywords: Socio-economic vulnerability, Physical-environmental vulnerability, GIS-MCDA, Random Forest, Flood vulnerability

Introduction

In the present century, floods are one of the most devastating and costly hazards worldwide, its contribution can be attributed to factors like climate change impacts on the hydrologic cycle, land-use changes and increased density of habitation activities in flood-prone areas. To mitigate the adverse impacts of flood, the design and implementation of risk management strategies are necessary. Since floods are common and cannot be prevented, flood risk assessment plays an important role in flood disaster management and decision making (Jiang et al. 2009). Risk is the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period, which is determined probabilistically (UN-ISDR 2009). It is estimated as the function of hazard, exposure and vulnerability (Corominas et al. 2014). The concept of vulnerability has different ideological and disciplinary perspectives which may vary with the theme, discipline, organization, researchers and stakeholders. (Dewan 2013). United Nations Office of Disaster Risk Reduction defines vulnerability as the conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards (UN-ISDR 2009). These factors have been detailed and studied by Stefania Balica and Nigel G. Wright in 2010, where the authors used many indicators to develop a methodology for flood vulnerability index (Balica and Wright 2010).

* Correspondence: deepak41194@gmail.com
Department of Civil Engineering, College of Engineering Trivandrum,
Thiruvananthapuram, Kerala, India

The perspective of vulnerability in the present study is similar to the Stefania Balica and Nigel G. Wright (2010). Here flood vulnerability is used as a multi-dimensional measure of the potential damage the human settlement inhabiting in the local self-government may encounter during a flood event, which is a totality of socio-economic conditions of the people and physical-environmental condition of the area. The Social and economic components of vulnerability comprise of factors such as the health of people, housing and living standards, education, social equity issues, income and wealth of the region (Rana and Routray 2017). The physical and environmental component represents the changes in milieu due to human or natural activities, topography, meteorological factors, and hydrologic parameters, which influence the vulnerability of the region to flood (Balica and Wright 2010).

In another aspect, the vulnerability can be found as a function of three related concepts such as exposure, susceptibility and resilience (Balica and Wright 2010). Susceptibility refers to the tendency of an area to undergo the effects of a specific hazardous process without taking into account either the moment of occurrence or potential victims and economic losses (Domínguez-Cuesta 2013). Resilience is the ability of a system, community or society exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner, through the preservation and restoration of its essential basic structures and functions through risk management (UN-ISDR 2009). Since vulnerability is geographically and socially distinguished, direct measurement of vulnerability is difficult (Adger 2006). In the present study, the term socio-economic vulnerability (SV) is used to indicate the capability of individuals and social systems to cope with flood hazards. This term encapsulates the social component and economic components of flood vulnerability. In the other aspect, exposure and resilience consideration to flood hazard is incorporated in this term. Physical and Environmental components of flood vulnerability are included in the term Physical-Environmental Vulnerability (PV), which considers the susceptibility aspect to hazard. Also, it can be observed that an area which may be classified as highly vulnerable under the socio-economic aspect may not be subjected to a higher degree of flood vulnerability under the physical-environmental domain. In the flood mitigation aspect, such over-estimations have to be minimized. The present study emphasizes on both physical-environmental and socio-economic aspects of vulnerability that can effectively be combined to get a more reliable flood vulnerability map spatially.

Remote sensing and Geographic Information System (GIS) techniques have been successfully made use of in flood vulnerability assessment (Khosravi et al. 2016; Dewan et al. 2007; Haq et al. 2012). Also a variety of approaches

such as GIS overlay analysis (Periyasamy et al. 2018), Multi-criteria decision analysis (Paulo et al. 2015; El Moussaoui et al. 2017), fuzzy approach (Lee et al. 2015), Random forest approach (Lee et al. 2017) etc. have been used in vulnerability related studies. Geographic Information System (GIS) based Multi-criteria decision analysis (MCDA) is a collection of methods for evaluating and combining geographic data and user preferences to aid in decision making (Malczewski 2006). Because of its flexible nature for analysing multiple and conflicting criteria, MCDA approach can be widely used to spatially represent flood vulnerability related studies (Giupponi et al. 2014, De Brito and Evers 2016, Feizizadeh and Kienberger 2017). Moreover, interests from different stakeholders are incorporated in MCDA in the form of criteria weights, thereby providing a solution for a multi-dimensional problem such as flood vulnerability which could be viewed under economic, physical, social, environmental perspectives (Tsoutsos et al. 2009). Analytical Hierarchy Process (AHP) is a widely used tool for multi-criteria decision making, providing an effective means to address complex problems in decision making (Vijith and Dodge-Wan 2019). Random forest is a machine learning algorithm that has given higher accuracy in feature classification problems (Breiman 2001; Cutler et al. 2007). A combination of multiple decision trees was used and the prediction result of these trees was ensembled to get the final prediction. Each tree is formed from certain samples by bootstrapping from the whole dataset. The advantage of random forest method includes: the model can accommodate the effect of missing data; it gives accurate prediction than decision trees; less affected by over-fitting problems; gives better error measurements; has a statistical advantage since randomness (bootstrapping and variable selection) is incorporated at multiple stages, gives better results when training data is noisy, estimates the contribution of each variable to output result (Solomon and Liu 2010; Dong et al. 2013; Chen and Ishwaran 2012). The ability of Random forest to determine the contribution of each parameter is very beneficial for better analysis of event and decision making, thereby making it highly suitable for analysis of classification problems.

Thus, the vulnerability in the present study is quantified, identifying various indices under the two main categories of socio-economic and physical- environmental vulnerability. Vulnerability in physical-environmental facet for a particular location is the same for all people in that location. In contrast, the vulnerability in socio-economic aspects for a particular location can vary within the group of people residing in that area, depending on their adaptive capacity. The primary emphasis of the study is to illustrate a methodology by which socio-economic vulnerability (SV) and physical-environmental (PV) can be made use in the assessment of vulnerability to floods in an area, using geospatial platforms, to

generate the vulnerability map that can be made use of by local self-governments as a critical tool in flood mitigation measures. The study makes use of MCDA approach for assessment of socio-economic vulnerability and Random forest method for the assessment of physical-environmental vulnerability and the two were integrated to generate the final vulnerability distribution map of the study area. The methodology has been applied to the vulnerability assessment of the Aluva Municipality (local self-government) in Ernakulam district, Kerala state, India. The details pertaining to the methodology used in the study, analysis of the problem to the study area, along with the results are discussed subsequently.

Methodology

The methodology was framed to assess the vulnerability of the study area. Vulnerability is susceptibility to suffer losses; in other words, weakened resilience to face the onslaught of a disaster. It incorporates considerations of both the intrinsic value of the elements concerned and their functional value in contributing to communal well-being in general and to emergency response and post-disaster recovery in particular. Socio-economic vulnerability is owing to adverse social positioning due to BKW000172poverty, unemployment, living in hazard prone zones, or dilapidated structures. On the other hand, the physical-environmental vulnerability refers to the influence of topography, hydrologic, and environmental parameters associated with flood propagation.

The vulnerability assessment of floods for a region thus encompasses two phases. In phase one, the creation of socio-economic vulnerability was done. For a group of people within an area of identical physical environmental conditions, these factors can individually contribute to their vulnerability to flood hazard. The contribution of these factors are subjective and is analysed based on expert opinion, and hence multi-criteria decision analysis approach is used. In Phase two, preparation of the physical-environmental vulnerability map was done. Independent of the social and economic patterns of the habitation, this vulnerability varies over the study area, depending on the topography and conditions of the terrain. Random forest, which is a highly efficient machine learning classification algorithm, is used to categorize the physical-environmental vulnerability.

The resulting spatial representations of the two phases were subjected to weighted overlay analysis to generate the vulnerability classification map following the procedure shown in Fig. 1.
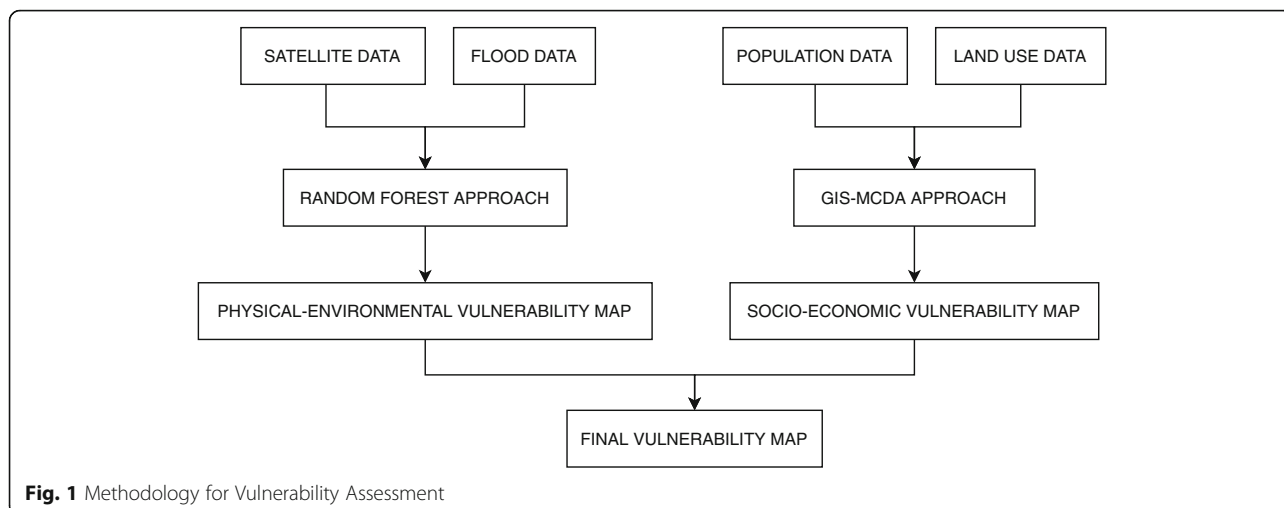
Socio-economic vulnerability

Indices such as age group, gender, number of members in a family, the function of buildings, type of roof covering, condition of buildings, unemployment and literacy rate of people, occupancy of people, distribution of population density, building density over the area, and land use or land cover classification of the area were considered to estimate socio-economic vulnerability. It is harder for children, women and aged people to cope with a flood event and are considered more vulnerable. Further, families having a higher number of individuals are also more vulnerable when their mobility during evacuation and coping capacity are taken into account. In case of the type of buildings, a commercial building is expected to contain a lesser population during a flood event and is less prioritized. Buildings that are well built (concrete houses) and are in good condition with good roofing are considered to be less vulnerable since they have more ability to withstand the effects of flood and heavy rains. Unemployed people have to depend on others for financial aids and will have to bear more economic burdens during floods and are considered more vulnerable under the economic perspective. Similarly, illiterates are often uneducated and unaware of disaster preparedness and management, hence more vulnerable. Furthermore, the area with a higher population density and building density is more exposed to flood hazards. And in the exposure viewpoint, urban area and agricultural area are more prioritized compare to barren lands in the assessment.

The following procedure was adopted for analysing the socio-economic vulnerability. First, a hierarchical structure was created using population data and land use data by analysing various factors that influence the vulnerability of the region. The population data which was obtained from census department of India includes age, gender, number of members in a family, function of the building, roof types of buildings, condition of buildings, unemployment, occupancy (Landlord/ Tenant), literacy rate, number of buildings and population density. This was used along with land use data derived by supervised classification to form the four objectives namely population, buildings, economics and exposed elements. These were considered as the three-level hierarchical classification model (Paulo et al. 2015; Kirby et al. 2019). ArcGIS software was used to create the different map layers for the various parameters considered in the study. The next step involves the rescaling of each of the factors into a linear scale of 1 to 10 and relative weights were assigned to each of the factors using AHP (Saaty 1977). These different factors were combined using a weighted linear combination which is a simple additive weighing procedure according to the following equation:

$$F(x) = \sum W_i . X_i \qquad (1)$$

W_i = normalised criteria scores

Fig. 1 Methodology for Vulnerability Assessment

Xi = criteria weights

The aggregation method of weighted linear combination (WLC) was used to map the socio-economic vulnerability and to classify the region into high, medium and low vulnerable areas of flood hazard. Since socio-economic vulnerability relates to the adaptive capacity of the population to that hazard, an area can be considered highly vulnerable, if the population within the area has less capacity to resist the impact of the natural hazard and to recover from its long term or short term effects. The socio-economic vulnerability can be considered as a subjective term. For example, commercial buildings are highly vulnerable in case of an economic point of view, while they are less vulnerable in case of population point of view. This makes it necessary to evaluate the socio-economic vulnerability under four different perspectives. The schematic workflow of the approach is presented in Fig. 2.

Physical-environmental vulnerability

In the present study, Elevation, Proximity to the river, Slope, Normalized Differential Vegetation Index (NDVI), Land use/land cover patterns (LULC), Stream Power Index (SPI), and Topographic Wetness Index (TWI) are the factors identified to affect the physical-environmental vulnerability (Haghizadeh et al. 2017, Samanta et al. 2018). Flooding in the study area is due to heavy rains and associated overflow of water from the river channel to nearby areas. As the nearness to the river increases the vulnerability to flooding event also increases. Moreover, a higher elevation and slope posses lesser chances of holding the excess water causing the flooding and consequently decreases the vulnerability. TWI and SPI are two topographic indices used that influence the flood (Moore et al. 1991; Pourghasemi et al. 2013). An increase in TWI and SPI over an area increases the flood vulnerability of the area to flood events. NDVI is a measure of vegetation cover over the area.

Interception losses in tree cover and infiltration of water to earth surface increases with NDVI and thereby make the area less vulnerable. (Wang et al. 2003). Vulnerability in the physical-environmental aspect is further analysed by the landuse landcover pattern of the area. Builtup area and roads cannot percolate water to subsurface and can make the area more vulnerable whereas paddy fields in the area can enhance the penetration of water to subsurface layers of earth and thus decreasing chance of flooding.

For creating a spatial map of physical-environmental vulnerability using Random forest model, a set of raster layers were prepared as in Fig. 3. The number of layers depends on the number of parameters that are considered as hazard inducing factors for the flood event. All indices considered in the study were derived from satellite data and made into raster format as layers in GIS platforms. The following equations were used (Wang et al. 2003).

$$NDVI = \frac{NIR - R}{NIR + R} \qquad (2)$$

(Where NIR = Near Infra-Red band and R = Red band) (Moore et al. 1991; Pourghasemi et al. 2013)

$$TWI = \frac{a}{\tan b} \qquad (3)$$

$$SPI = a \times \tan b \qquad (4)$$

(Where a = upslope area and $\tan b$ = local slope in radians)

The dataset was prepared by stacking these raster layers and this served as the input to the Random forest model. Random forest model randomly samples a certain number of values by bootstrapping from the different parameter values of the dataset and divides them

Fig. 2 Methodology for Socio-economic vulnerability
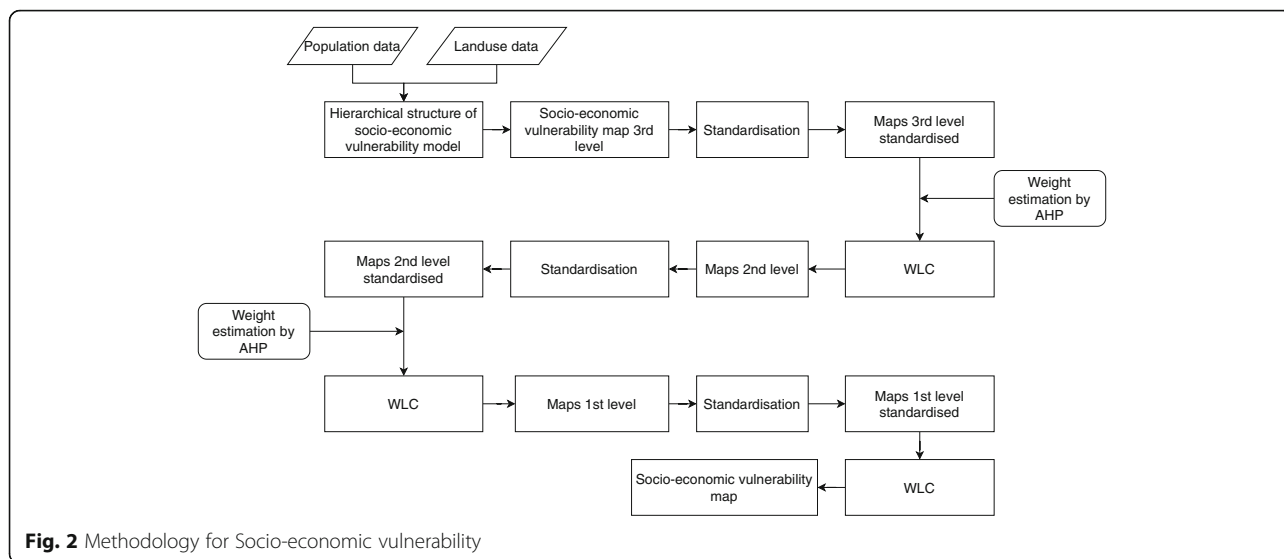
into training and testing samples. A number of decision trees were developed to form the random forest and the end node of these trees was indexed as high, medium or low-vulnerable based on training data created. Training data was another raster layer created using known flood water levels at different locations over the study area and interpolating the same to get a raster layer. To minimize the uncertainties due to interpolation of flood water levels, the surface elevation data of the area was also incorporated with the water level to get the required training raster. The pixels of the training raster have values which correspondingly represent the three vulnerable zones considered in the study. The model can thus

be trained to three different vulnerability zones namely high, medium and low. The Random Forest model used the same set of classification trees formed to classify the pixels of the testing sample set and the model accuracy was also checked. When the required accuracy level was met with, the whole data set of the study area was fed to the RF model prepared and the model classify every pixel into any of the three categories thereby into the three-vulnerable zone namely high, medium and low-vulnerable to the flood hazard. A highly physical-environmental vulnerable area is those which are at higher flood risk due to its physical and environmental characteristics. For instance, area at a higher elevation is

Fig. 3 Methodology for physical-environmental vulnerability
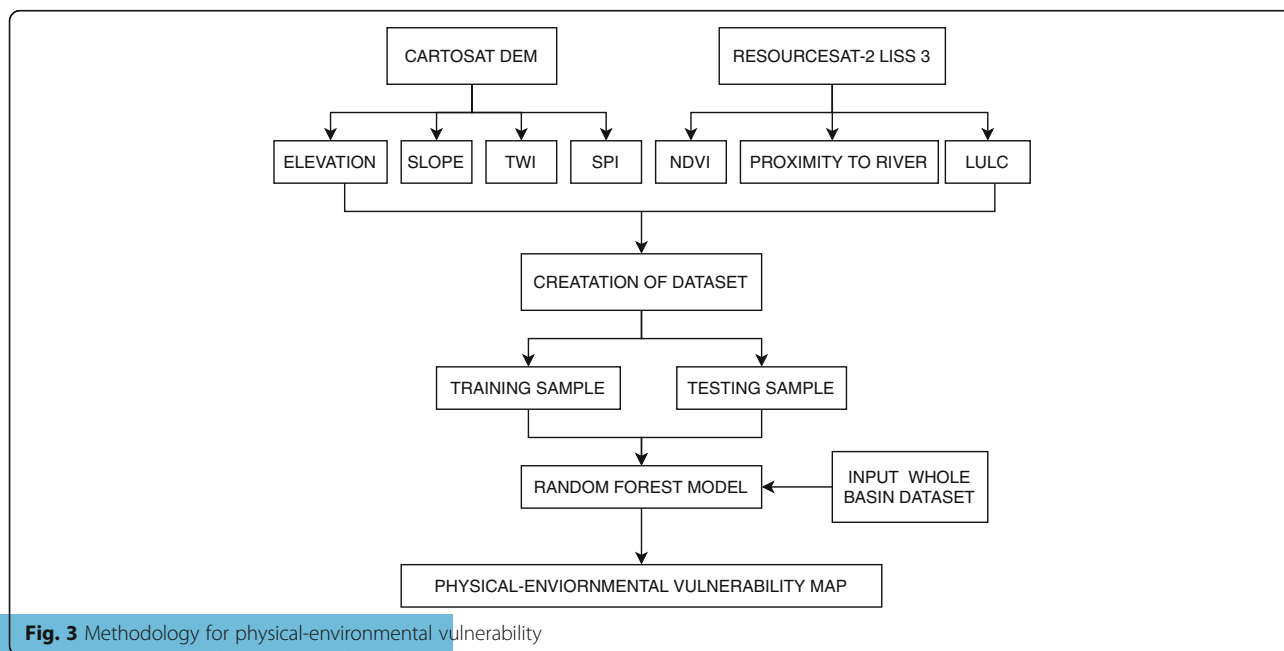
less vulnerable, while the area at closer proximity to the river is more vulnerable. A combined assessment of such physical and environmental factors is made use of to classify the study area to different vulnerable zones. A low vulnerable zone is less susceptible to flood and can be considered suitable for human settlements. The algorithm was implemented in python platform.

Combined vulnerability to floods

A region can be considered vulnerable to flood if it is both socio-economic and physical-environmental vulnerable. This concept can prevent over estimation in many cases as a socio-economic vulnerable region may not be physically-environmental vulnerable. Thus, the socio-economic vulnerability (SV) obtained from MCDA approach and the physical-environmental vulnerability (PV) from Random Forest method was combined using AND operator to get the spatial distribution of vulnerable zones (V) within the study area. This can be expressed as:

$$V = SV \cap PV \tag{5}$$

An analysis was also performed by varying the weights of socio-economic and physical-environmental vulnerabilities, to study the effect of variation of socio-economic and physical-environmental vulnerability in the distribution of vulnerable zones.

Analysis

The details pertaining to the study area, data used in the study, assessments of SV and PV for the study area are discussed subsequently.

Study area

The proposed methodology was illustrated for the Aluva town, a peri-urban municipality in the northern suburb of the city of Kochi, in Ernakulam district, Kerala state, India (Fig. 4). The town centre is located at latitude of 10.1004° N and longitude 76.3570° E, with the Periyar River flowing through the municipality, such that it almost divides the region into two. It is a town of around 25,000 residents with an area of 6.46 km^2 comprising of 23 wards as per 2011 census. Aluva town was the most affected municipality in Ernakulam district, during the 2018 Kerala flood. The influence of dams and other hydraulic structures which regulate the flood were not considered in this study.

Data used

The data used in the study include spatial data products like Cartosat DEM, satellite imagery, ward map of the municipality, Google image and the Inundation map with flood levels. The non-spatial data include population data and associated statistical data. The details pertaining to the data used in the study are presented in Table 1.
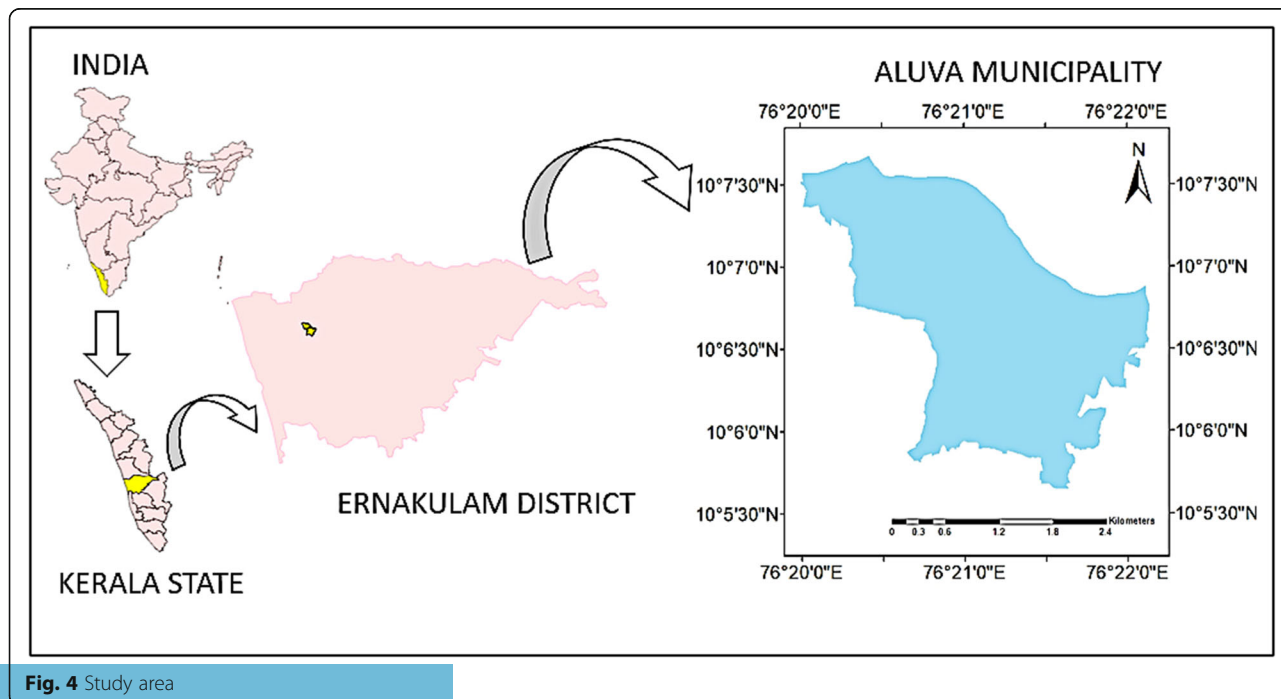

Fig. 4 Study area

Table 1 Data used in the study

SL NO	DATA	SOURCE	REMARKS
1	DEM	National Remote Sensing Centre (NRSC)	• Cartosat 1
			• Spatial Resolution: 1 arc sec
			• File format: GeoTIFF
2	LISS III	National Remote Sensing Centre (NRSC)	• Resourcesat 1
			• Spatial resolution: 23.5 m
			• File format: GeoTIFF
			• Number of bands: 4(2,3,4,5)
			B3: 0.62–0.68 (RED)
			B4: 0.77–0.86 (NIR)
3	Ward Map	Aluva Municipality	Hardcopy
5	Population data	2011Census data	Datasheets
6	Flood level data and Inundation map	Kerala State Disaster Management Authority (KSDMA)	Shapefiles

Analysis of socio-economic vulnerability

The hierarchical structure of socio-economic vulnerability model is shown in Fig. 5. The objective population considers age, gender, and the number of members per house as the second level factors. The third level classification for age was included as, age below 14 years, age between 14 and 65 years and age above 65 years, as the extremes of the age spectrum are more vulnerable to the disaster. In the case of gender, third level classification include male or female and for the number in family representing houses with members less than 4 or greater than 5, considering the fact that larger families would have more dependent to evacuate and thus have to share the resources.

The second level factors for the objective buildings include function, type of roof, and the condition of the buildings. Based on the type of use, buildings were classified into residential and commercial buildings. Based on the roof type, houses were classified into three as houses with concrete roof, houses with tiled roof, houses with other types of roofs, which include thatched houses, houses made of bamboo, slate etc. Since concrete houses can withstand the effect of

flooding, they were given comparatively less weightage and houses with other types of roofing were given higher weightage. A similar classification was followed for the condition of buildings. The buildings were classified into three as, houses with good living condition, liveable condition and with the dilapidated condition, with maximum weightage given to houses with dilapidate condition and minimum weightage to those with good condition.

The next objective of economics, considered unemployment, housing occupancy (Landlord/Tenant) and illiteracy as the factors. The unemployed are dependent on other family members and are considered more vulnerable. While considering housing occupancy, tenants usually do not possess the financial status to own a house and are thus economically more vulnerable. Illiterates were considered as more vulnerable as they generally lack the basic knowledge to adapt in a hazard situation.

While considering the objective of exposed elements, the second level factors include land use, population density, and building density. Under land use, built-up, agricultural area and barren land were considered as the
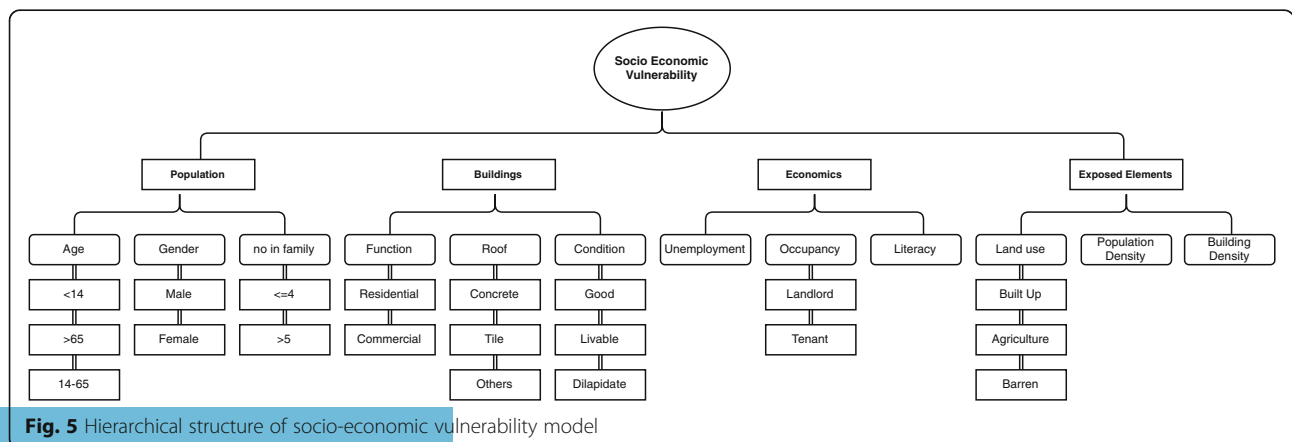

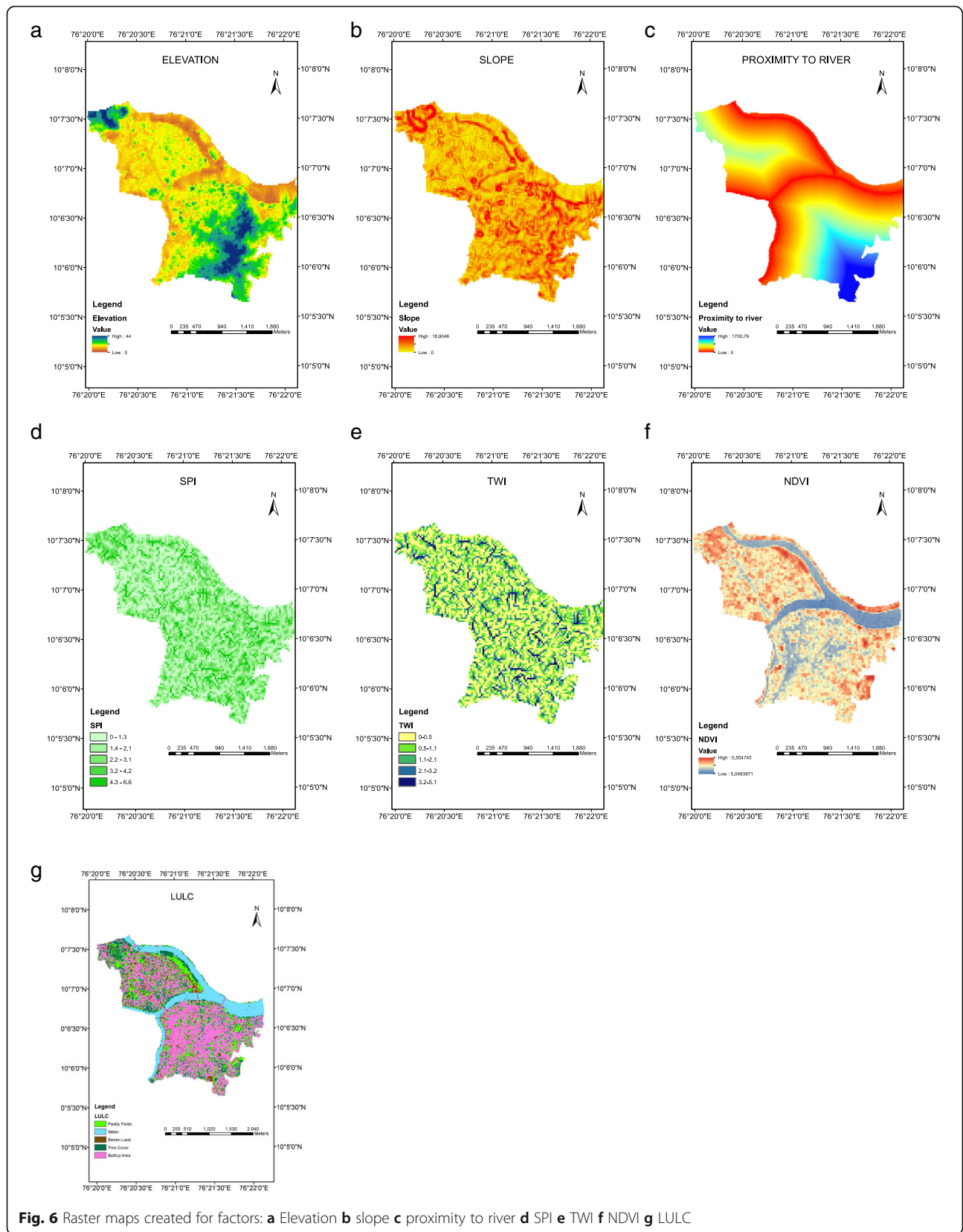**Fig. 5** Hierarchical structure of socio-economic vulnerability model

Fig. 6 Raster maps created for factors: **a** Elevation **b** slope **c** proximity to river **d** SPI **e** TWI **f** NDVI **g** LULC

third level factors. The urban area was considered as highly vulnerable to a flood event; whereas the vulnerability of agricultural land is dependent upon on the crop season and the barren land always possess relatively low vulnerability. The population density and density of buildings are also important factors that influence socio-economic vulnerability in risk areas.

According to the hierarchical classification, three-level of maps were prepared. The process of map preparation assumes negligible effect of the floating population within the study area. Various map layers were created in the ArcGIS platform. The higher-level maps were prepared by overlaying the lower level map of the corresponding factor. To avoid the sharp variation of population and building densities over the ward boundaries, the population and building density maps were created using pycnophylatic interpolation by employing focal statistics tool in ArcGIS.

The weights were assigned to each factor using Analytical Hierarchy process (AHP). After creating the hierarchical structure of socio economic vulnerability model, the relative importance of each pair of criteria of the model is evaluated. Saaty's 9- point continuous scale (Saaty 1980) was used for weighing each criterion for the creation of a pairwise comparison matrix. First the pairwise comparison matrix was created using each of the attributes within the third level factors. The score of one indicates that both criteria were equally important whereas a score of 9 indicates extreme importance of one criterion over the other. All scores were assembled in a pairwise comparison matrix with the value of 1 on the diagonal and reciprocal scores on the lower left triangle. The pairwise comparison generated was based on expert opinion which was evaluated. An eigen vector was extracted from each comparison matrix and the weight was assigned to each of the factors (Leal 2020) within the third level. For each level in hierarchy it is necessary to understand the consistency of judgement to accept the results of judgement. The parameter consistency ratio (CR) is used to check the consistency and the value of CR greater than 0.1 has to be re-evaluated. By trial and error, the weights for each factor in the third level were obtained. Using each of these weights, weighted overlay was performed on the third level factors to create the second level factors. Similarly, weight estimation and overlay for each of the second level and third level factors were performed respectively to obtain the final vulnerability map.

The measurement of each of these factors was done on different scales. Thus, rescaling of these factors to a common scale was necessary. The factors were standardised to a linear scale of 1 to 10, where 1 represents very low vulnerability and 10 represents very high vulnerability. A simple additive weighting procedure such as weighted linear combination was used to combine each criterion. The 3rd level maps were combined using WLC to create 2nd level maps and further these 2nd level maps were clustered to generate 1st level maps based on the four above mentioned objectives. The final social vulnerability map was obtained by combining the four objectives by assigning equal weightage to them.

Table 2 Weight for Socio-economic vulnerability factors

Sl no	Function	Weight	Sl no	Function	Weight
1	Age	0.668		*Other roofs*	0.655
	Age less than 14	0.473	6	Condition of house	0.387
	Age greater than 65	0.473		*Good*	0.068
	Age between 14 to 65	0.052		*Livable*	0.249
2	Gender	0.088		*Dilapidated*	0.681
	Male	0.25	7	Employment	0.694
	Female	0.75	8	Occupancy	0.179
3	Persons per house	0.243		*Landlords*	0.16
	Members Less than 5	0.125		*Tenants*	0.833
	Members Greater than 4	0.875	9	Illiterates	0.126
4	Building function	0.443	10	Land use	0.088
	Residential	0.125		*Built up*	0.681
	Commercial	0.875		*Agriculture*	0.249
5	Roof type of buildings	0.169		*Barren*	0.068
	Concrete roof	0.054	11	Population density	0.668
	Tiled Roof	0.289	12	Building density	0.243

Analysis of factors affecting PV

Raster layers were created using ArcGIS and Erdas Imagine software from remote sensing data for the analysis and then used to create a dataset to be fed into the Random forest model. Elevation of the area from the mean sea level was one of the factors affecting the physical-environmental vulnerability. A Digital Elevation Model (DEM) was used to get the spatial variation of elevation in the study area. The Cartosat DEM of one arc second spatial resolution (approximately 30 m) was used in this study. Those pixels which lie in higher elevation are considered to be less prone to flood as compared to those in low-lying areas. The slope was another aspect considered in the study, derived from DEM using GIS tools, with the index reflecting the degree of topographic change. Regions with high slopes drain the water faster compared to the flat regions that are more susceptible to flooding. Proximity to the river was another most important factor affecting the vulnerability. It is considered that an area having closer proximity to the river is more
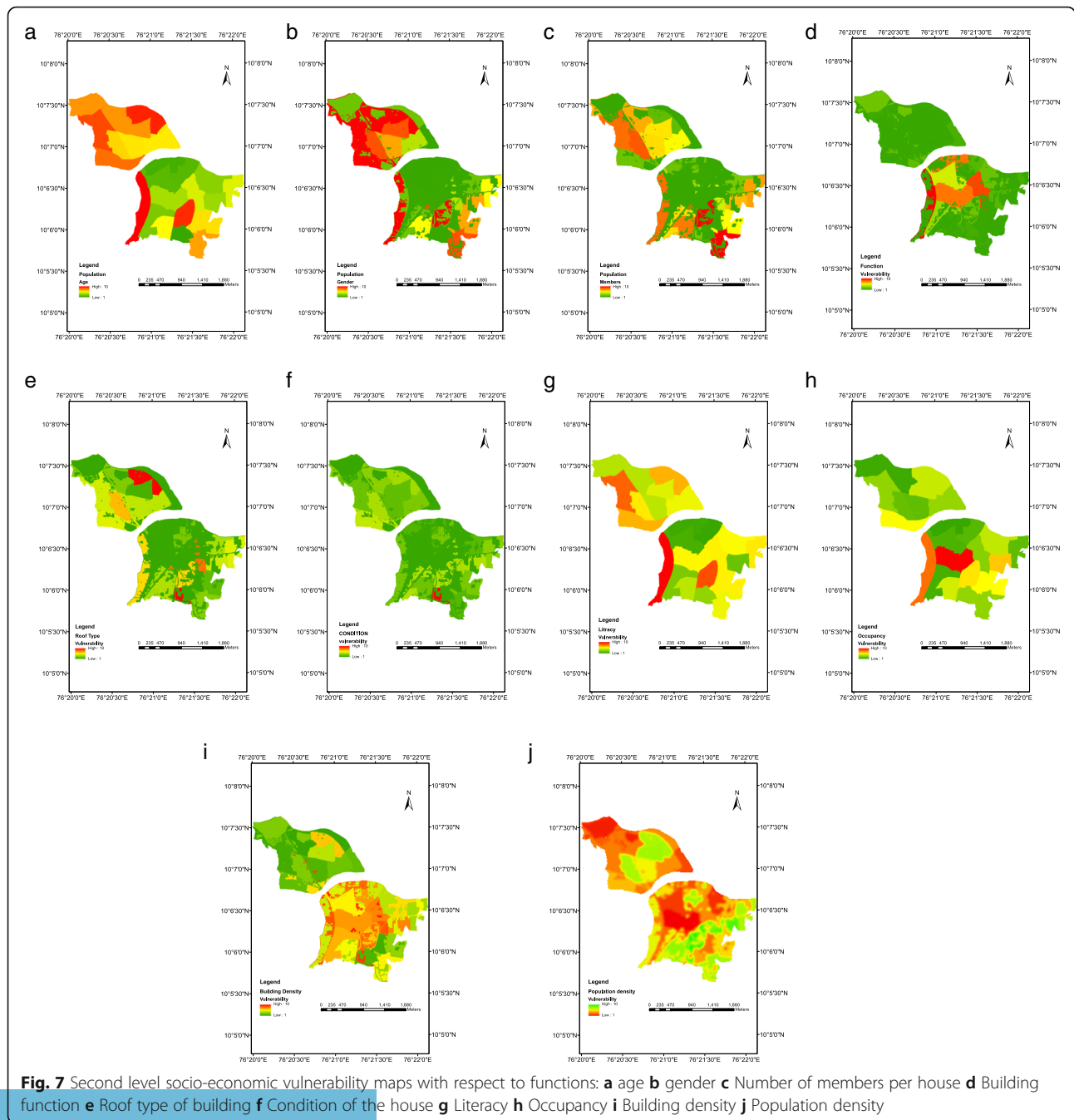

Fig. 7 Second level socio-economic vulnerability maps with respect to functions: **a** age **b** gender **c** Number of members per house **d** Building function **e** Roof type of building **f** Condition of the house **g** Literacy **h** Occupancy **i** Building density **j** Population density

Deepak *et al. Geoenvironmental Disasters*   (2020) 7:35

Page 11 of 19

likely to be affected by the flooding. Since both proximity to river and elevation are incorporated in the study, a combination of situations like an area nearer to the river but having higher elevation was considered as less vulnerable. Proximity to the river was estimated by computing the Euclidean distance of every point on the area considered concerning the digitized line feature (from LISS 3 image) representing the river. In addition, two major topographic indices considered were topographic wetness index (TWI) and stream power index (SPI). Both are derivatives of DEM and computed using the formula given in eq. 3 and eq. 4. The topographic wetness index, also known as the compound topographic index (CTI), is a steady-state wetness index. TWI was determined as the product of slope and upstream contributing area per unit width perpendicular to the flow direction. TWI has a good correlation with many soil properties such as horizon depth, percentage of silt, organic matter and phosphorous and hence directly influence the flood vulnerability (Moore et al. 1991). Potential flow erosion at a particular point is represented by stream power index. Erosion risk increases as the amount and velocity of water increases and hence directly contributes to flood vulnerability (Pourghasemi et al. 2013). Normalized Difference Vegetation Index (NDVI) was used to quantify the vegetation cover of an
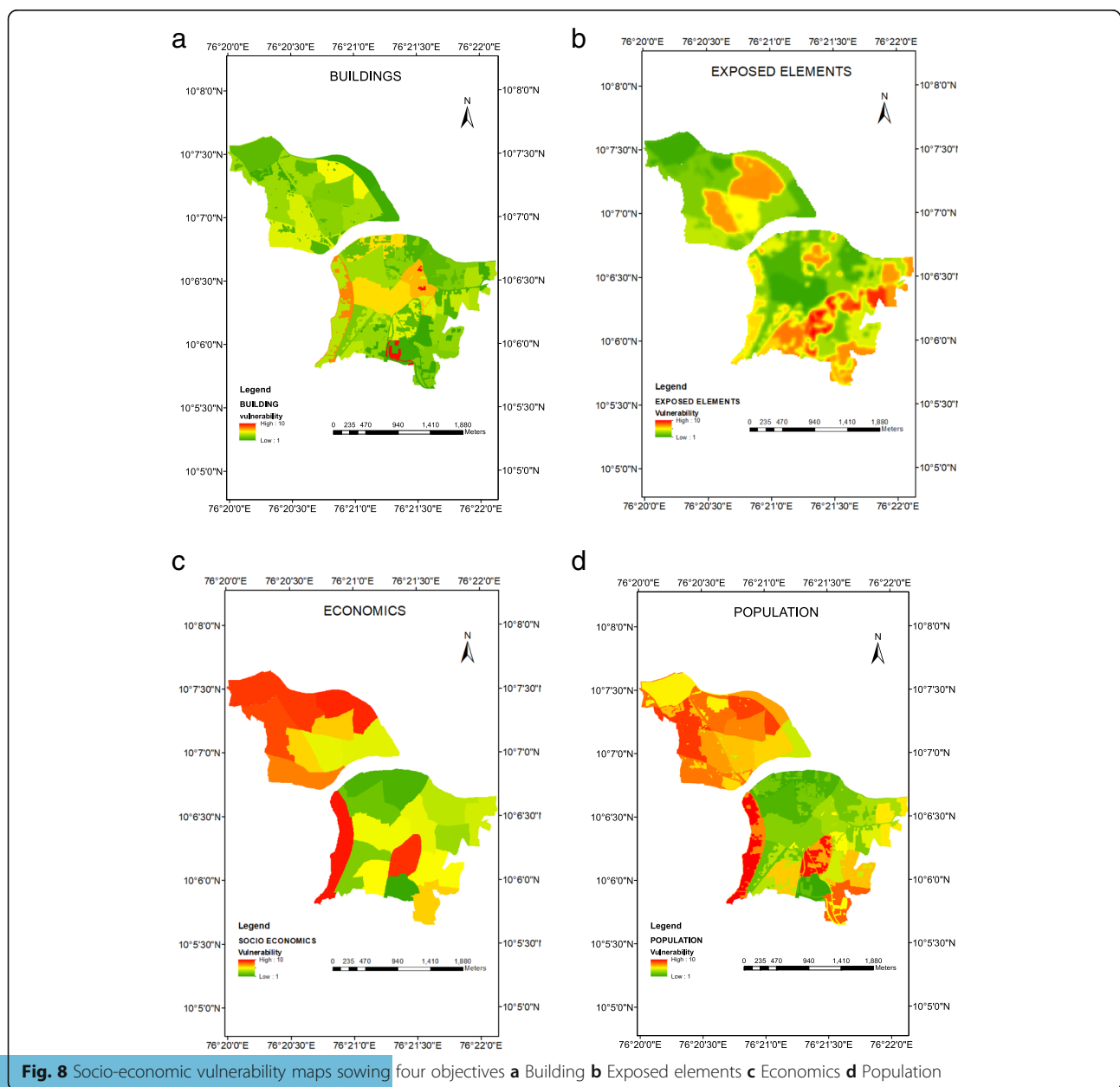

Fig. 8 Socio-geoenomic vulnerability maps sowing four objectives **a** Building **b** Exposed elements **c** Economics **d** Population

Deepak *et al. Geoenvironmental Disasters*   (2020) 7:35

Page 12 of 19

area. It was computed using eq. 2 from LISS III image using band 3 and band 4, being the Red and Near Infrared bands. Increase in NDVI shows thicker vegetation cover which reduces the runoff. Interception loss to surface runoff increases with vegetation and the vegetation facilitates more infiltration which in turn reduces the surface runoff causing the flood. Least NDVI values are usually shown by water bodies and barren lands where the uninterrupted flow of water can occur with very less penetration to the subsurface (Wang et al. 2003). Hence a decrease in NDVI was considered to be an aid to increase flood hazard vulnerability. Land Use/ Land Cover of the area is also a key factor which determines the intensity of flood. Five land classes

over the study area were identified by field studies and using the supervised classification of LISS 3 image, the study area was classified into Built-up area, Water cover, Tree-cover, Barren land and Paddy fields. Built-up area includes regions where buildings and roads exist, with least percolation of water to the sub-surface; whereas paddy fields are recharge zones where water easily reaches the subsurface layers. Thereby built-up area contributes to flood hazard vulnerability, whereas paddy fields and barren lands can decrease the chances of the area being flooded. Different raster maps created for all the indices considered to be hazard inducing factors were analysed. All the maps were made into the same resolution by using nearest neighbour resampling technique. The
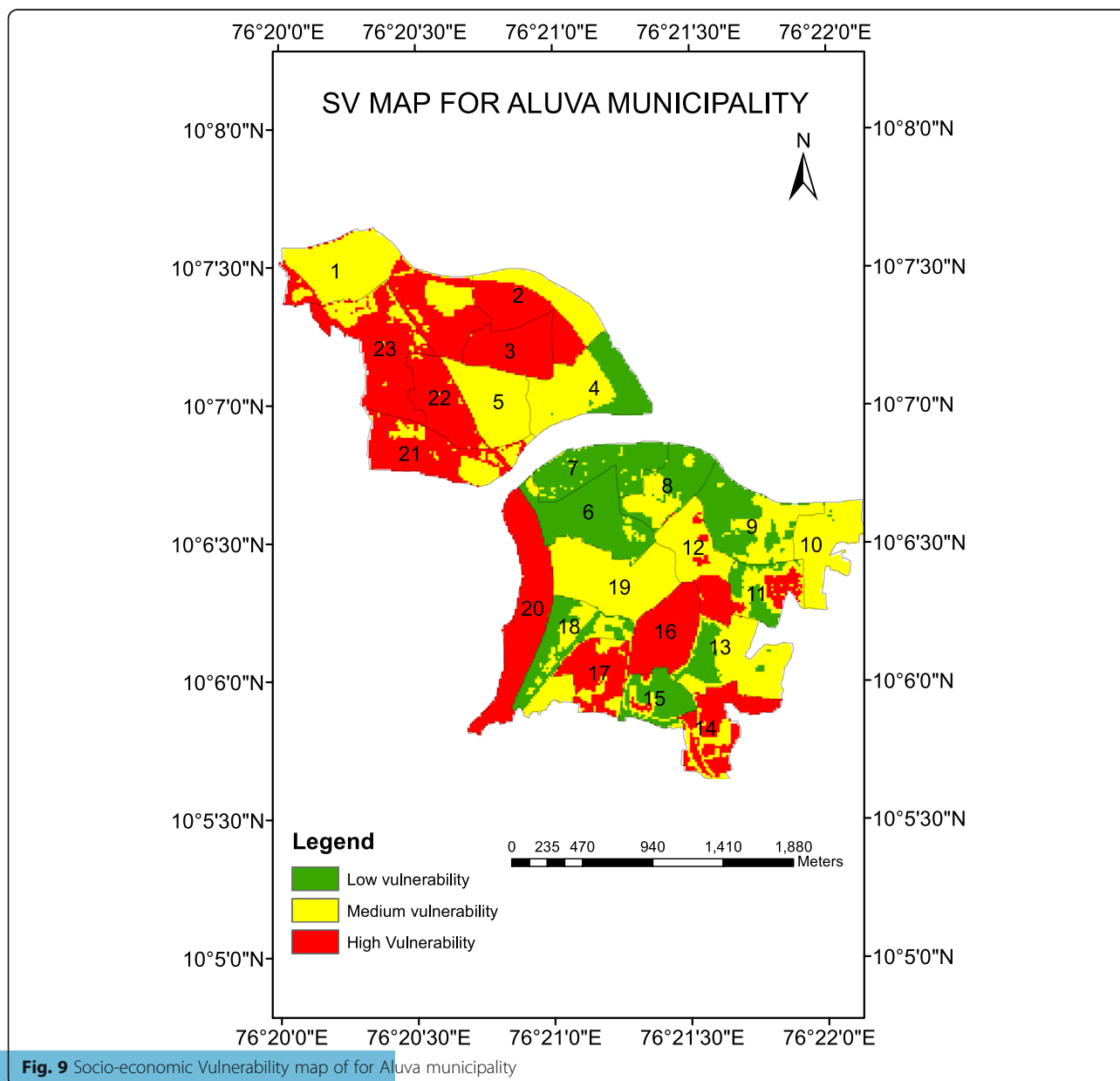

Fig. 9 Socio-Geoenomic Vulnerability map of for Aluva municipality

output maps with the spatial distribution and intensity variations across the Aluva municipality as in Fig. 6 was obtained.

Analysis of random forest model

Spatial variation of all the physical-environmental factors influencing the flood vulnerability over the study area was plotted into raster layers in ArcGIS, which in turn form the data set as input to the Random forest model. The variables in the model are the physical-environmental factors of which some are randomly selected by bootstrapping and decision trees are grown on these selected samples identified as the training dataset. A number of decision trees are randomly grown on these datasets for which the end nodes are labelled as High, Medium and Low vulnerable based on the real flood data. The algorithm for forming random forest implemented in Python is summarised as follows:

1. Samples were selected from the dataset by bootstrapping procedure. (Yeh et al. 2014)
2. Best split on randomly selected variables was found based on Gini value and trees are

grown based on training data (Wang et al. 2015)

3. Data to be predicted was given to the model and final prediction was done by ensembling the output of each decision tree through majority voting principle.

The accuracy of the model was obtained using testing samples. The model was run on testing samples and the predicted output of testing samples was compared with the known values to get the accuracy. The influence of parameters considered was also determined by the model. This was done using a function called Mean Gini decrease index which computes Gini decreases individually for all parameters over all the trees of the forest (Ai et al. 2014). In addition, the sensitivity of the model to the number of trees grown and to the depth of pruning was also assessed. Accuracy of the model was also checked for different combinations of tree number and pruning depth. The predicted output was visualized in ArcGIS software. The ward map of the Aluva municipality was superimposed over the vulnerability map to get the ward wise distribution of the physical-environmental vulnerability.

Table 3 Ward-wise distribution of socio-economic vulnerability for Aluva municipality

Ward NO.	High vulnerability (%)	Medium vulnerability (%)	Low vulnerability (%)
1	41.02	57.02	1.96
2	59.99	39.94	0.07
3	99.05	0.95	0
4	0.2	44.9	54.91
5	0.75	96.04	3.21
6	0.13	0.09	99.78
7	0	0	100
8	0.2	0.48	99.33
9	0	0.48	99.51
10	0	77.19	27.31
11	27.01	25.71	47.27
12	33.16	66.09	57.16
13	1.16	61.05	37.79
14	74.64	16.77	8.55
15	6.54	18.35	76.11
16	98.89	0.64	0.47
17	77.08	3.35	19.56
18	0.29	9.46	90.25
19	0.11	10.51	89.37
20	99.7	0.02	0.28
21	72.3	27.7	0
22	73.6	26.4	0
23	72.5	27.5	0

Results and discussions

The study illustrates the assessment of the socio-economic vulnerability and physical-environmental vulnerability in generating the vulnerability to the flood at the local self-government level. The vulnerability to flood greatly depends on the physical-environmental factors are further modified by combing the socio-economic vulnerability, so as to get a realistic picture of the spatial variation of vulnerability. The summary of the outcomes of the study is as follows.

Socio-economic vulnerability

The standardised criteria weights for different parameters obtained using AHP for the study area are presented in Table 2. The second level maps produced for the different functions from the third level maps are shown in Fig. 7. The above second level maps were aggregated using the weighted linear combination to obtain the socio-economic vulnerability maps with respect to four objectives, namely population, buildings, economics and exposed elements to produce the first level

maps as shown in Fig. 8. In the socio-economic vulnerability assessment of the population in the municipality, an approximately equal percentage of low vulnerable area and high vulnerable area are seen in the study. High vulnerable areas are seen mostly in the residential areas within the municipality. In the case of socio-economic vulnerability with respect to exposed elements and buildings, the municipality has a medium vulnerability. Some portions of ward no. Fifteen exhibited very high vulnerability with respect to these objectives, which must be probably due to the presence of thatched houses along with houses having dilapidated living conditions in the ward. Also, the areas that are considered vulnerable with respect to the objective buildings exhibit a low to medium vulnerability with respect to the objective exposed elements. This clearly identifies that building density and characteristics of buildings are two distinct parameters which must be considered separately. With respect to objective of economics, the municipality exhibited high vulnerability.
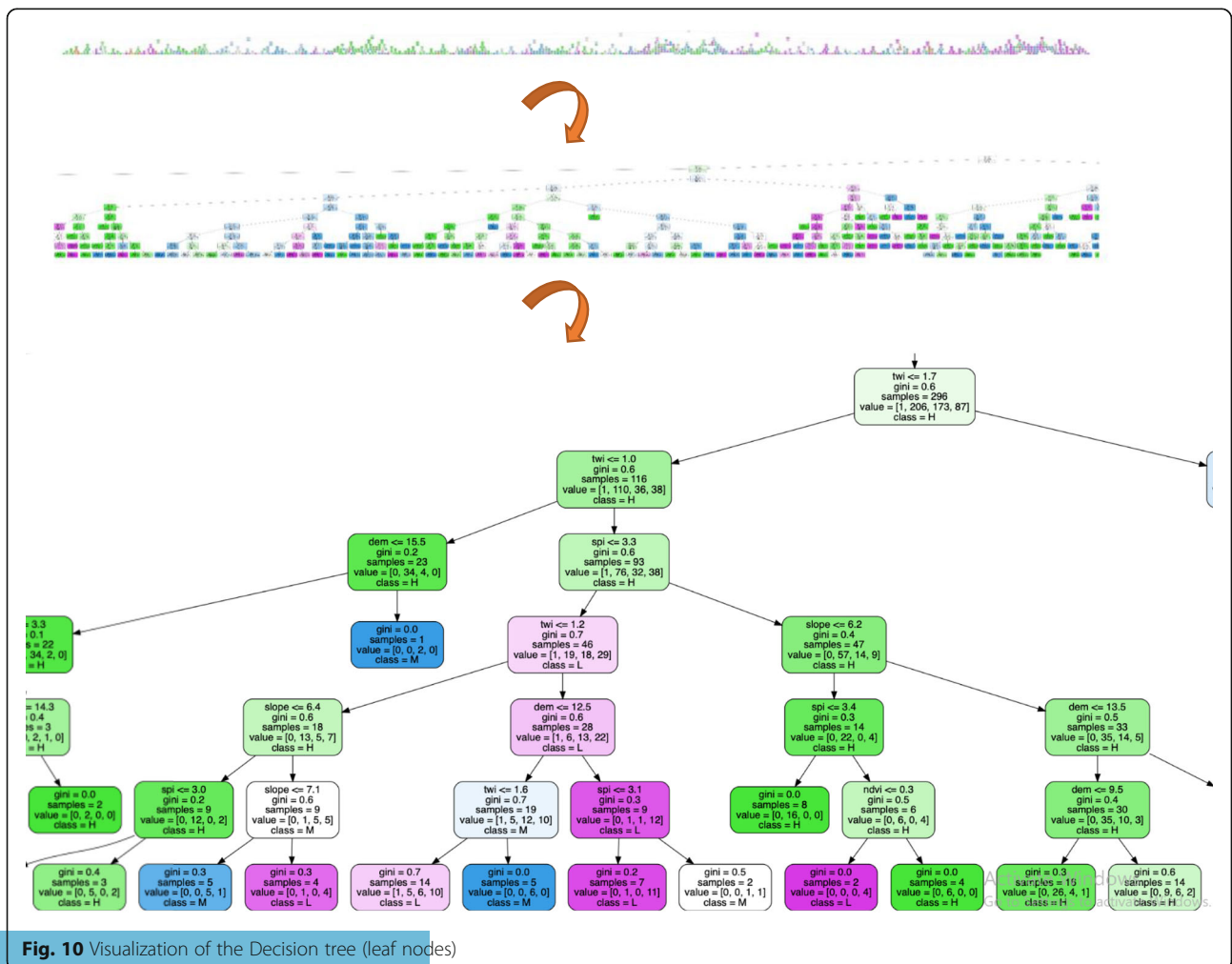

Fig. 10 Visualization of the Decision tree (leaf nodes)

Deepak *et al. Geoenvironmental Disasters*   (2020) 7:35

Page 15 of 19

The four objectives associated with socio-economic vulnerability (Fig. 8), were given equal weightage and combined using WLC operator to create the socio-economic vulnerability map for Aluva municipality as in Fig. 9. The percentage of the area coming under high, medium, and low vulnerability zones in each ward are tabulated in Table 3. With respect to socio-economics, the municipality exhibited high vulnerability. About 44.13% of the total area was under high vulnerable zones and requires proper planning strategies. From the table, it is clear that ward numbers 3, 16, and 20 exhibits the most socio-economic vulnerable area, with a score greater than 98% of its area under the high-vulnerability zone. This was followed by wards 14, 17, 21, 22, and 23

with more than 70% of the area falling under high socio-economic vulnerability zones. A low socio-economic vulnerability was being exhibited by ward no: 6, 7, 8, 9, 15, 18 and 19 probably because of the smaller number of households.

Physical-environmental vulnerability

The Random forest model was formed using algorithm coded in Python and the following analysis was done. The nodes of one such decision tree in the model analysis are as shown (Fig. 10).

Here the number of pixels classified according to particular pruning process is shown as value. The values within the parenthesis show the number of pixels that
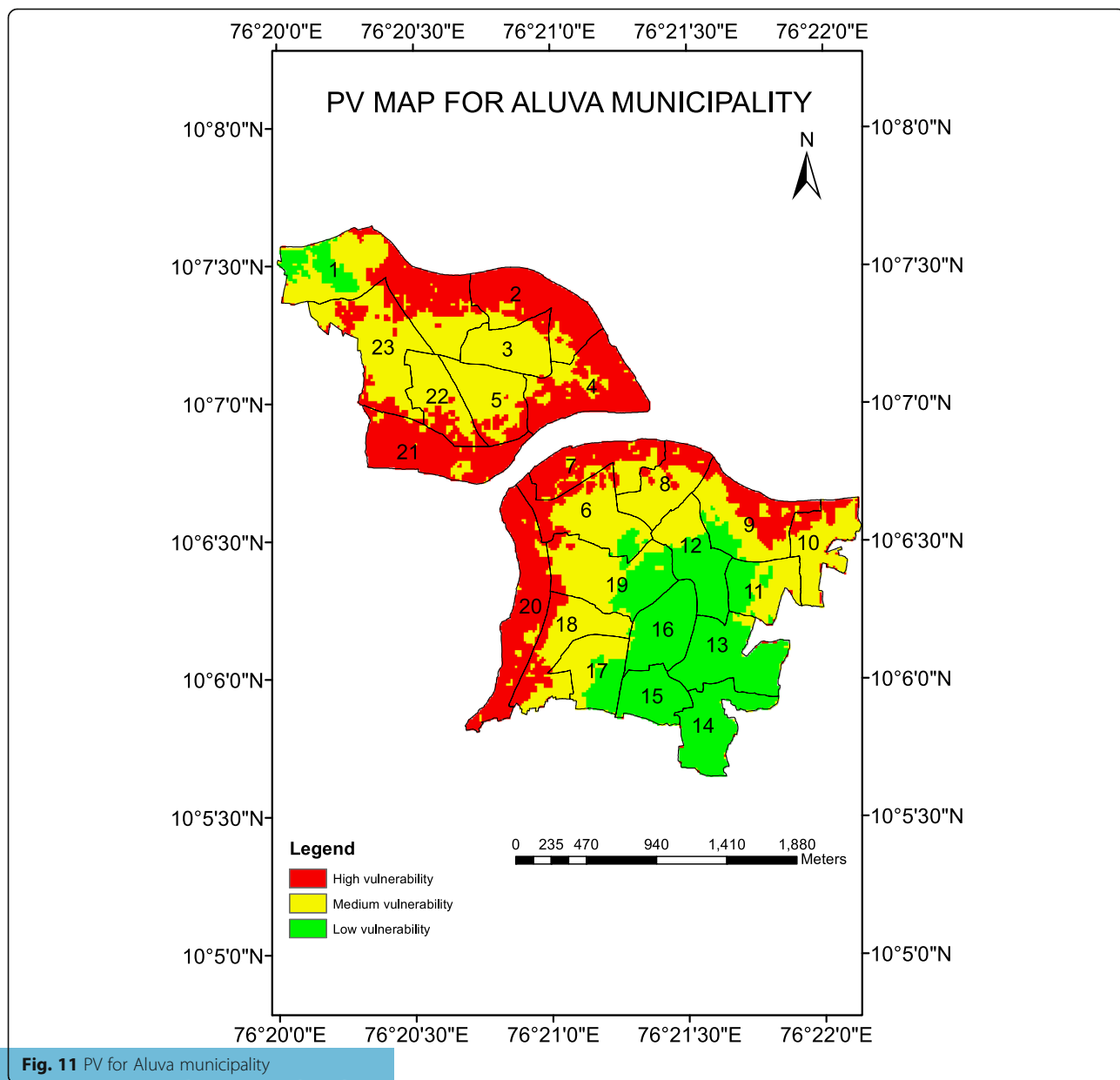

Fig. 11 PV for Aluva municipality

Deepak *et al. Geoenvironmental Disasters*   (2020) 7:35

Page 16 of 19

Table 4 Ward-wise distribution of PV to flood in Aluva municipality

Ward NO.	High vulnerability (%)	Medium vulnerability (%)	Low vulnerability (%)
1	36.50	49.85	13.65
2	83.54	16.46	0.00
3	9.97	90.03	0.00
4	85.19	14.81	0.00
5	24.36	75.64	0.00
6	28.54	64.58	6.88
7	75.65	24.35	0.00
8	35.03	64.97	0.00
9	44.12	45.70	10.18
10	24.53	75.47	0.00
11	1.06	65.20	33.74
12	0.00	34.95	65.05
13	0.50	5.39	94.11
14	0.59	1.91	97.51
15	0.08	1.42	98.50
16	0.00	2.29	97.71
17	0.27	62.98	36.75
18	31.34	68.56	0.11
19	7.58	59.76	32.66
20	96.22	3.78	0.00
21	95.23	4.77	0.00
22	26.20	73.80	0.00
23	25.79	74.21	0.00
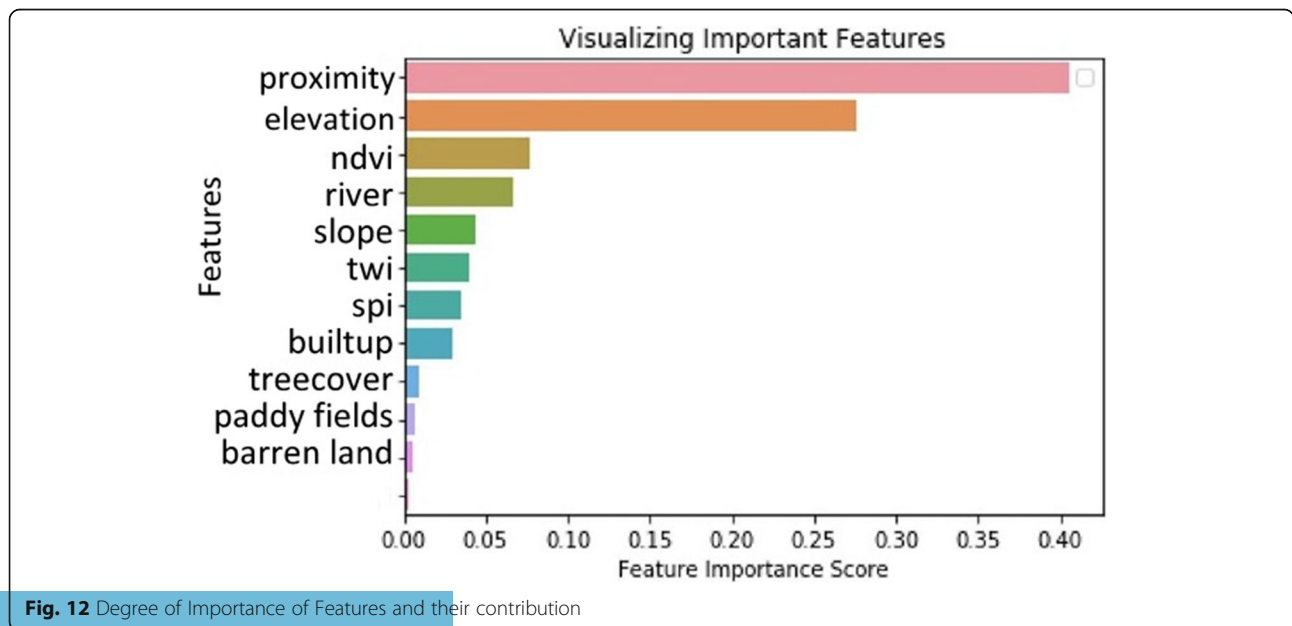
Fig. 12 Degree of Importance of Features and their contribution

belongs to a particular class and the end results was given by the tree as H, M and L, indicating Highly vulnerable, Medium vulnerable and Low vulnerable pixels respectively. After building the model using training samples, the model formed was run on testing samples to check the accuracy. The predicted output was compared with the known values in testing samples. The accuracy of the model by running on the testing sample was 98.79% which is acceptable. The output predicted was obtained in .csv format. This was converted to vector point data and imported into ArcGIS software and interpolated to get PV map. The ward map was then placed over the above vulnerability map to identify the spatial distribution of the vulnerability zones across different wards, as in Fig. 11. The area-wise distribution of

vulnerability zones over different wards are tabulated in Table 4.

From the table, it can be seen that ward numbers 21, 4, 2, 20 and 7 are classified under high PV of which ward 20 is at the highest vulnerability of 96%. Wards 15, 16, 14, 13 and 12 have a majority of their area under low risk, of which ward 15 is least vulnerable with 98% of its whole area under the low vulnerable zone. The contributing features in PV for the study area derived from the model are shown in Fig. 12.

It is evident that the proximity or distance from river and elevation are the two major contributors of physical vulnerability in the study area, and thereby the most affecting factors of flood hazard risk when compared to the other factors, as seen in Fig. 12. From feature

Fig. 13 Vulnerability map using different combinations of PV and SV **a** 0.5 PV-0.5 SV, **b** 0.6 PV-0.4 SV **c** 0.7 PV-0.3 SV **d** 0.8 PV-0.2SV **e** 0.9 PV-0.1 SV

Deepak *et al. Geoenvironmental Disasters*   (2020) 7:35

Page 18 of 19
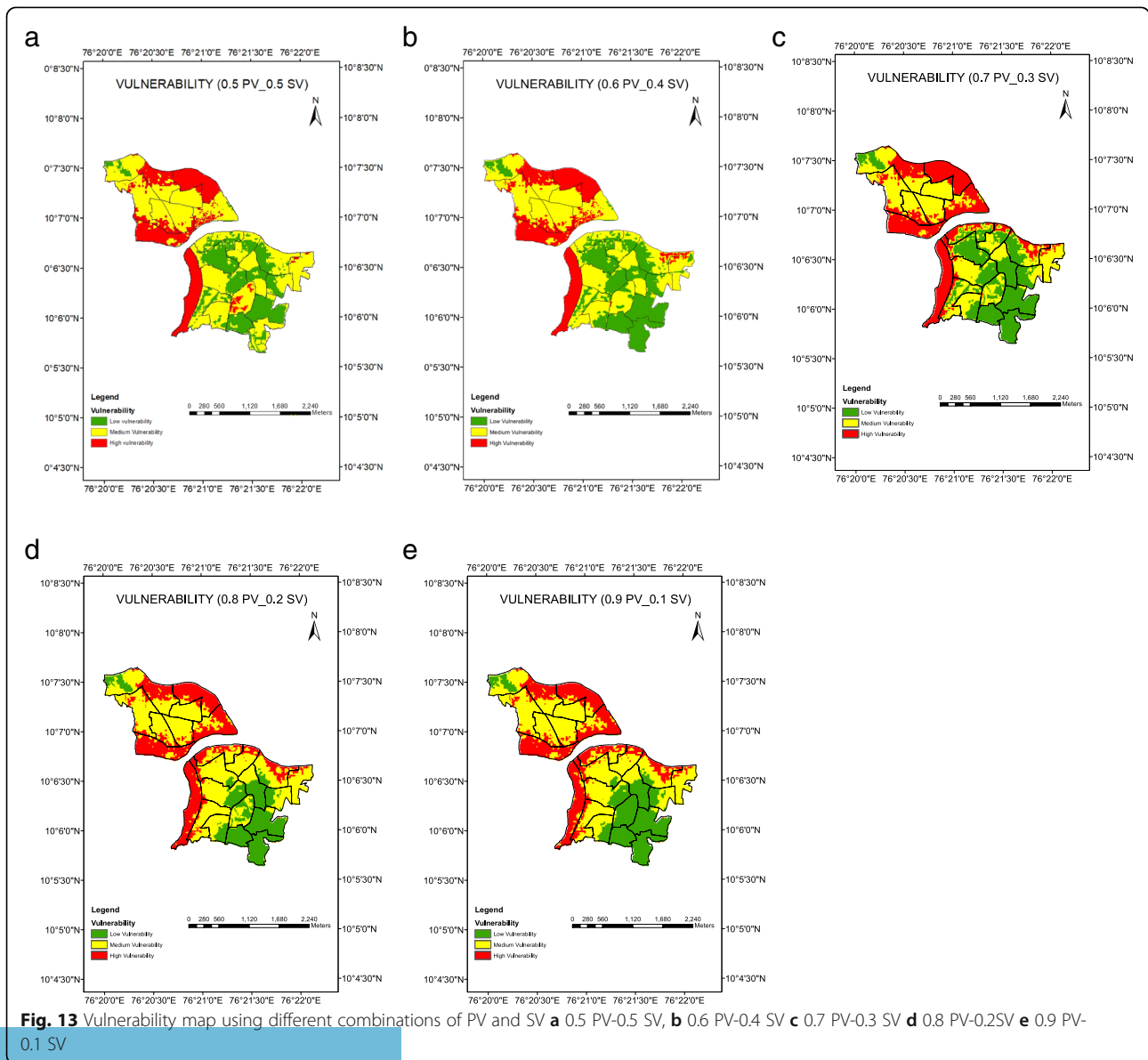
importance score in figure, it can be seen that proximity to river followed by elevation are the two factors in determining the vulnerability in the case studied. In the study area considered, only some area having higher elevation were nearer to river, and most of highly elevated area were farther from river. This particular physiography of the region must have resulted in more importance to proximity than the elevation. Among the land use consideration, the built-up area has a higher score when compared to other forms of land use, as the resistance to flow is much higher in a built–up area thereby increasing the severity of flooding.

Vulnerability map

The overall vulnerability map of Aluva municipality (local self-government) was obtained by combining the socio-economic and physical-environmental vulnerability using weighted averaging. Different combinations were done for physical and socio-economic vulnerability. During events of large floods (with a high return period of the order of 1000 years or more), the whole area will be submerged. In such a case, the physical-environmental vulnerability does not have relevance for decision-makers, but socio-economic vulnerability can play a key role to assess the resilience of the community and for taking further measures. Whereas in cases of seasonal floods during usual monsoon rainfall (with low values of return period), the physical-environmental vulnerability has more importance since the area of impact will be based on the physical and environmental factors considered than the social and economic conditions of the area. A combination of socio-economic and physical-environmental vulnerability can give a realistic picture of the vulnerability to flood in the study area, to help in the computation of flood risk. Thus, the weightage of each vulnerability depends on the forecasted flood hazard event and the priorities of stakeholders. In the present study, different cases were done by varying the weights to the PV and SV, and the overall vulnerability to flood for the study area was generated. The combinations of vulnerability maps prepared with weightage for PV in the range 0.5 to 0.9 with weightage for SV in the range 0.1 to 0.5 are illustrated in Fig. 13. The decision-maker (local self-government) can select the appropriate vulnerability map based on the weightage to be given for the socio-economic and physical-environmental aspects of vulnerability, as illustrated in the Figure.

Conclusion

The focus of the study lies in the assessment of vulnerability to floods by combing the four major criteria of vulnerability into two aspects, namely the physical-environmental vulnerability and socio-economic vulnerability to floods of an area. The methodology was applied to the Aluva municipality (local self-government) in Ernakulam district, Kerala state, India. The geospatial approach was made use of in the preparation of vulnerability maps. The contribution of various factors associated with socio-economic and physical-environmental vulnerability assessment was also done. Following the approach, the vulnerable population can be identified along with their location and thus proper mitigation measures can be adopted in a necessity. When the combinations of socio-economic and physical-environmental vulnerability were considered there was a significant variation in the spatial distribution of various vulnerability zones based on the weightage given for PV and SV. These maps can be effectively used depending on the intensity of hazard forecasted based on which the respective authority can take appropriate decisions. Thus the proposed method is more effective since the over-estimation of the vulnerability of areas which are highly socio-economic vulnerable but not physical-environmental vulnerable can be reduced.

References
Adger WN (2006) Vulnerability. Glob Environ Chang 16(3):268–281
Ai FF, Bin J, Zhang ZM, Huang JH, Wang JB, Liang YZ, Yu L, Yang ZY (2014) Application of random forests to select premium quality vegetable oils by their fatty acid composition. Food Chem 143:472–478. https://doi.org/10.1016/j.foodchem.2013.08.013
Balica S, Wright NG (2010) Reducing the complexity of the flood vulnerability index. Environ Hazards 9(4):321–339. https://doi.org/10.3763/ehaz.2010.0043
Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
Chen X, Ishwaran H (2012) Random forests for genomic data analysis. Genomics 99(6):323–329. https://doi.org/10.1016/j.ygeno.2012.04.003
Corominas J, van Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S et al (2014) Recommendations for the quantitative analysis of landslide risk. Bull Eng Geol Environ 73(2):209–263. https://doi.org/10.1007/s10064-013-0538-8

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

De Brito MM, Evers M (2016) Multi-criteria decision making for flood risk management: a survey of the current state-of-the-art. Nat Hazards Earth Syst Sci Discuss 3:6689–6726. https://doi.org/10.5194/nhessd-3-6689-2015

Dewan AM (2013) Hazards, risk, and vulnerability. In: Floods in a Megacity. Springer, Dordrecht, pp 35–74. https://doi.org/10.1007/978-94-007-5875-9c https://doi.org/10.1016/j.gloenvcha.2006.02.006

Dewan AM, Islam MM, Kumamoto T, Nishigaki M (2007) Evaluating flood hazard for land-use planning in greater Dhaka of Bangladesh using remote sensing and GIS techniques. Water Resour Manag 21(9):1601–1612. https://doi.org/10.1007/s11269-006-9116-1

Domínguez-Cuesta MJ (2013) Susceptibility. In: Bobrowsky PT (ed) Encyclopedia of natural hazards, Encyclopedia of Earth Sciences Series. Dordrecht, Springer. https://doi.org/10.1007/978-1-4020-4399-4

Dong LJ, Li XB, Kang PENG (2013) Prediction of rockburst classification using random Forest. Trans Nonferrous Metals Soc China 23(2):472–477 https://doi.org/10.1016/S1003-6326(13)62487-5

El Moussaoui S, Omira R, Zaghloul MN, El Talibi H, Aboumaria K (2017) Tsunami hazard and buildings vulnerability along the Northern Atlantic coast of Morocco–the 1755-like tsunami in Asilah test-site. Geoenviron Disasters 4(1): 25. https://doi.org/10.1186/s40677-017-0089-6

Feizizadeh B, Kienberger S (2017) Spatially explicit sensitivity and uncertainty analysis for multicriteria-based vulnerability assessment. J Environ Plan Manag 60(11):2013–2035. https://doi.org/10.1080/09640568.2016.1269643

Giupponi C, Mojtahed V, Gain A, Balbi S, Biscaro C (2014) An integrated approach for including social capacities, and economic valuation in risk assessment of water related hazards in uncertain scenarios. In: Proceedings of the 7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA https://ssrn.com/abstract=2455821

Haghizadeh A, Siahkamari S, Haghiabi AH et al (2017) Forecasting flood-prone areas using Shannon's entropy model. J Earth Syst Sci 126:39. https://doi.org/10.1007/s12040-017-0819-x

Haq M, Akhtar M, Muhammad S, Paras S, Rahmatullah J (2012) Techniques of remote sensing and GIS for flood monitoring and damage assessment: a case study of Sindh province, Pakistan. Egypt J Remote Sens Space Sci 15(2): 135–141. https://doi.org/10.1016/j.ejrs.2012.07.002

Jiang W, Deng L, Chen L, Wu J, Li J (2009) Risk assessment and validation of flood disaster based on fuzzy mathematics. Prog Nat Sci 19(10):1419–1425. https://doi.org/10.1016/j.pnsc.2008.12.010

Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Nat Hazards 83(2):947–987. https://doi.org/10.1007/s11069-016-2357-2

Kirby RH, Reams MA, Lam NS, Zou L, Dekker GG, Fundter DQP (2019) Assessing social vulnerability to flood hazards in the Dutch Province of Zeeland. Int J Disaster Risk Sci 10(2):233–243

Leal JE (2020) AHP-express: a simplified version of the analytical hierarchy process method. MethodsX 7:100748. https://doi.org/10.1016/j.mex.2019.11.021

Lee MJ, Kang JE, Kim G (2015) Application of fuzzy combination operators to flood vulnerability assessments in Seoul, Korea. Geocarto Int 30(9):1052–1075. https://doi.org/10.1080/10106049.2015.1027290

Lee S, Kim JC, Jung HS, Lee MJ, Lee S (2017) Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics Natural Hazards Risk 8(2):1185–1203. https://doi.org/10.1080/19475705.2017.1308971

Malczewski J (2006) Ordered weighted averaging with fuzzy quantifiers: GIS-based multi-criteria evaluation for land-use suitability analysis. Int J Appl Earth Obs Geoinf 8(4):270–277. https://doi.org/10.1016/j.jag.2006.01.003

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5(1):3–30. https://doi.org/10.1002/hyp.3360050103

Paulo F, Mourato S, Moreira M (2015) Social vulnerability assessment of flood risk using GIS-based multi criteria decision analysis. A case study of Vil Nova de Gaia (Portugal). Geomatics Natural Hazards Risk. https://doi.org/10.1080/19475705.2015.1052021

Periyasamy P, Yagoub MM, Sudalaimuthu M (2018) Flood vulnerable zones in the rural blocks of Thiruvallur district, South India. Geoenviron Disasters 5(1):21. https://doi.org/10.1186/s40677-018-0113-5

Pourghasemi HR, Jirandeh AG, Pradhan B, Xu C, Gokceoglu C (2013) Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J Earth Syst Sci 122(2):349–369. https://doi.org/10.1007/s12040-013-0282-2

Rana IA, Routray JK (2017) Socioeconomic vulnerability assessment: a case study of flood prone urban communities of Pakistan. In: International expert forum: mainstreaming resilience and disaster risk reduction in education. Springer, Singapore, pp 123–139

Saaty T (1977) A scaling method for priorities in hierarchical structures. J Math Psychol 15(3):234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Saaty T (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, New York

Samanta S, Pal DK, Palsamanta B (2018) Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. Appl Water Sci 8:66. https://doi.org/10.1007/s13201-018-0710-1

Solomon T, Liu Z (2010) Earthquake induced damage classification for reinforced concrete buildings. Struct Saf 32(2):154–164. https://doi.org/10.1016/j.strusafe.2009.10.002

Tsoutsos T, Drandaki M, Frantzeskaki N, Iosifidis E, Kiosses I (2009) Sustainable energy planning by using multi-criteria analysis application in the island of Crete. Energy Policy 37(5):1587–1600. https://doi.org/10.1016/j.enpol.2008.12.011

UN-ISDR (2009) Terminology of disaster risk reduction. United Nations International Strategy for Disaster Reduction, Geneva https://www.undrr.org/terminology

Vijith H, Dodge-Wan D (2019) Modelling terrain erosion susceptibility of logged and regenerated forested region in northern Borneo through the Analytical Hierarchy Process (AHP) and GIS techniques. Geoenviron Disasters 6(1):8. https://doi.org/10.1186/s40677-019-0124-x

Wang Q, Watanabe M, Hayashi S, Murakami S (2003) Using NOAA AVHRR data to assess flood damage in China. Environ Monit Assess 82(2):119–148. https://doi.org/10.1023/A:1021898531229

Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X (2015) Flood hazard risk assessment model based on random forest. J Hydrol 527:1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

Yeh CC, Chi DJ, Lin YR (2014) Going-concern prediction using hybrid random forests and rough set approach. Inf Sci 254:98–110. https://doi.org/10.1016/j.ins.2013.07.011

Publisher's Note