

# Geospatial Approach For Assessment Of Vulnerability To Flood In Local Self Governments

Deepak S (✉ [deepak41194@gmail.com](mailto:deepak41194@gmail.com))

College of Engineering Trivandrum <https://orcid.org/0000-0002-2305-2036>

Gopika Rajan

College of Engineering Trivandrum

Jairaj P G

College of Engineering Trivandrum

---

## Research

**Keywords:** Socio-economic vulnerability, Physiographic vulnerability, GIS-MCDA, Random Forest, Flood Vulnerability

**Posted Date:** April 17th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-22816/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on December 14th, 2020. See the published version at <https://doi.org/10.1186/s40677-020-00172-w>.

# Abstract

Recent years have shown a significant increase in the occurrence of flood globally, with 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 which combines vulnerability under physiographic 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 GIS tools. The machine learning algorithm, Random Forest, which uses hazard inducing factors as input was implemented for the evaluation of physiographic vulnerability. The social vulnerability of the region was analysed using GIS Multi-criteria decision analysis approach, 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.

## 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, it is necessary to determine the aerial extent of flood susceptibility and vulnerability to plan, design and implement management strategies by local governing bodies and other stakeholders as part of flood mitigation measures (Sarkar 2020). Vulnerability illustrates the condition of an object under consideration and represents its disadvantage to the impact of a hazard (Fekete 2009). It can be considered as the product of the interaction between environmental and social processes (Parker 2000). Recently remote sensing and Geographic Information System (GIS) techniques have been successfully made use of for flood vulnerability assessment (Khosravi et al. 2016; Dewan et al. 2007; Haq et al. 2012). A variety of approaches have been proposed to estimate vulnerability such as frequency ratio (Lee et al., 2012), Multi-criteria decision analysis (Paulo et al., 2015), fuzzy approach (Lee et al., 2015), Random forest approach (Lee et al., 2017) etc.

Usually, vulnerability to flood is determined based on the proximity of the area to the river. Also, it is observed in many studies, that other physiographic factors such as elevation, slope, landcover etc. contribute to the flood vulnerability in addition to proximity to river (Samanta et al., 2018). Moreover, the flood vulnerability is also associated with the vulnerability of community assessed based on the socio-economic condition of people in the area (Kirby et al., 2019). In contrast, we can observe 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 physiographic domain. In flood mitigation aspect such over-estimations have to be minimized. The present study emphasizes on both physiographic and socio-

economic aspects of vulnerability that can effectively be combined to get a more reliable flood vulnerability map spatially.

Geographic Information System (GIS) based Multi-criteria decision analysis (MCDA) is a collection of methods for evaluating and combining geographic data and user's preferences to aid in decision making (Malczewski et al., 2006). Because of its flexible nature for analysing multiple and conflicting criteria, MCDA approach can be widely used to spatially represent flood vulnerability (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 multidimensional problem such as flood vulnerability which could be viewed from different perspectives such as economic, physical, social, physiographic etc (Tsoutsos et al. 2009). These methods are effective in analysing the vulnerability but are subject to some uncertainties since the index weights in these methods are usually given manually and are therefore subjective (Stefanidis and Stathis, 2013).

Random forest is a machine learning algorithm which 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. With these advantages, the random forest approach was expected to be highly appropriate in classification problems.

The major emphasis of the study is to illustrate a methodology by which socio-economic vulnerability (SV) and physiographic vulnerability (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 key 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 physiographic 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 vulnerability assessment of floods for a region encompasses two phases: Phase 1: Creation of socio-economic vulnerability using multi-criteria decision analysis and Phase 2: Preparation of the physiographic vulnerability using the Random forest algorithm. The resulting spatial representations were subjected to weighted overlay analysis, to generate the vulnerability classification map following the procedure shown in Figure 1.

*Socio-economic vulnerability:* The following procedure was adopted for analysing the socio-economic vulnerability. First, a hierarchical structure was created by analysing various factors that influence the vulnerability of the region. A three-level hierarchical classification was adopted which comprises of four components namely: Population, Buildings, Socio-economics and Exposed elements (Paulo et al., 2015). Arc GIS software was used to create the different map layers for the various parameters considered for 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 analytic hierarchy process (AHP) developed by Saaty (1977). MCDA allows the decision-makers to compare different criteria based on their relative importance to obtain a final solution. These different factors were combined using a weighted linear combination which is a simple additive weighing procedure according to the following equation:

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

$W_i$  = normalised criteria scores

$X_i$  = 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 effects, whether long term or short term. 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. In that context, an area labelled as less vulnerable in socio-economic domain indicates that the population residing in that area are more resilient to the flood hazard, whereas a highly vulnerable area in the same domain represents that community would experience adverse impacts of flood hazard and can be regarded as less resilient. The schematic workflow of the approach is presented in Figure 2.

*Physiographic Vulnerability.* For creating a spatial map of physiographic vulnerability using Random forest model, a set of raster layers were prepared as in Figure 3. The number of layers depends on the number of parameters that are considered as hazard inducing factors for the flood event. 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 flood vulnerability and were derived from satellite data in raster format layers in GIS platforms. The following equations were used.

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

(Wang et al., 2003)

(Where  $NIR$  = Near Infra-Red band and  $R$  = Red band)

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

(Moore et al., 1991).

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

(Pourghasemi et al., 2013)

(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 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 floodwater 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, 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 and thus the area can be divided into the three-vulnerable zone namely high, medium and low-vulnerable to the flood hazard. A highly physiographic vulnerable area is those which are at higher flood risk due to its physiographic characteristics. For instance, area at a higher elevation is less vulnerable, while the area at closer proximity to the river is more vulnerable. A combined assessment of such physiographic factors is used 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 physiographic vulnerable. This concept can prevent overestimation in many cases as a socio-economic vulnerable (SV) region may not be physiographically vulnerable (PV). Thus, the socio-economic vulnerability obtained from MCDA and the physiographic vulnerability from Random Forest

can be combined using AND operator to get the spatial distribution of vulnerable zones (V) within the study area. This can be expressed as:

$$V = SV \cap PV \quad \dots (5)$$

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

## **ANALYSIS**

The details of the study area, data used in the study, assessment of socio-economic and physiographic vulnerability 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 (Figure 2). The town centre is located at a 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<sup>2</sup> 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 regulates 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.

Table 1: Data used in the study

SL NO	DATA	SOURCE	REMARKS
1	DEM	National Remote Sensing Centre (NRSC)	<ul style="list-style-type: none"> <li>• Cartosat 1</li> <li>• Spatial Resolution: 1 arc second</li> <li>• File format: GeoTIFF</li> </ul>
2	LISS III	National Remote Sensing Centre (NRSC)	<ul style="list-style-type: none"> <li>• Resourcesat 1</li> <li>• Spatial resolution: 23.5m</li> <li>• File format: GeoTIFF</li> <li>• Number of bands: 4(2,3,4,5)</li> </ul> <p style="text-align: center;">B3 :0.62-0.68 (RED)</p> <p style="text-align: center;">B4 :0.77-0.86 (NIR)</p>
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 Figure 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 included male and female and for the number of persons per household, houses with members less than 5 and greater than 4 were taken, 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 included the number of rooms, the type of use, the material of roof type and the condition of the buildings. The third level classification for each of the second level factors was as follows: The number of rooms was divided in the 3<sup>rd</sup> hierarchical order as two, houses with room lesser than 5 and houses with room greater than 4. Based on the type of use, buildings were classified into residential and non-residential buildings. Based on the roof type material houses were classified into 3, houses with concrete roof, houses with tiled roofs, houses with other types of roofs, which included thatched houses, houses made of bamboo, slate etc. Since concrete houses can withstand the effect of flooding, they are given comparatively less weightage and houses with other types of roofing are given higher weightage. A similar classification was been followed in the condition of buildings. The buildings were classified into three, 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, socioeconomics, considered of unemployment, housing occupancy (Landlord/Tenant) and illiteracy. 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 can also be considered 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 included land use, population density, and building density. Among them, land use, urban areas, agricultural area and barren land are considered as the third level factors. The urban area can be highly vulnerable to a flood event. The vulnerability of agricultural land depends upon the crop season and barren land always possesses relatively low vulnerability. The density of people and 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 the effect of the floating population within the study area was ignored. Various map layers were created in the ArcGIS platform. The higher-level maps were prepared by overlaying the corresponding lower levels of the map of corresponding factors. Analytical hierarchy process (AHP) was employed to assign the weightage to each criterion and aggregation was done by weighted linear combination. 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 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. The weights were assigned to each factor using AHP. The consistency of judgement was checked by calculating the consistency ratio (CR) and any value of CR greater than 0.1 was re-evaluated (Saaty 1980).

A simple additive weighting procedure such as weighted linear combination was used to combine each criterion. The 3<sup>rd</sup> level maps were combined using WLC to create 2<sup>nd</sup> level maps and further these 2<sup>nd</sup> level maps were clustered to generate 1<sup>st</sup> level maps based on the four above mentioned objectives. The final social vulnerability map was obtained by combining each of the four objectives by providing equal weightage.

### **Analysis of factors affecting the physiographic vulnerability**

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 physiographic 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 was considered that an area having closer proximity to the river was more 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 is 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 representing the river. In addition, two major topographic indices considered were topographic wetness index and stream power index. The topographic wetness index (TWI), 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 area. It was computed using equation 2 from LISS III image where band 3 and band 4 are respectively Red and Near Infra-red 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 was also a key factor which determines the intensity of flood. Built-up area is those places 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. Thus, 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 to obtain the following output maps with the spatial distribution and intensity variations across the Aluva municipality as in Figure 6.

### **Analysis of Random forest model**

Spatial variation of all the physiographic 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 physiographic 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 physiographic risk.

## Results And Discussions

The study illustrates the assessment of the socio-economic vulnerability and physiographic vulnerability in generating the vulnerability to the flood at the local self-government level. The vulnerability to flood greatly depends on the physiographic 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

Table 2: Weight for Socio-economic vulnerability factors

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

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 Figure 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, socioeconomics and exposed elements to produce the first level maps as shown in Figure 8. In the socio-economic vulnerability assessment for 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. 15 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 socioeconomics, the municipality exhibited high vulnerability.

The above four objectives were given equal weightage and combined using WLC operator to create the socio-economic vulnerability map for Aluva municipality as in Figure 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.

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

## Physiographic Vulnerability

The Random forest model was formed using algorithm coded in Python and the following analysis were done. The nodes of one such decision tree in the model analysis are as shown (Figure 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 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 was acceptable. The output predicted was obtained in .csv format. This was converted to vector point data and was imported into ArcGIS software and interpolated to get physiographic vulnerability 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 Figure 11. The area-wise distribution of vulnerability zones over different wards are tabulated in Table 4.

Table 4: Ward-wise distribution of Physiographic vulnerability 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

From the table, it can be seen that ward numbers 21, 4, 2, 20 and 7 are classified under high physiographic risk of which ward 20 was 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 was the least vulnerable with 98 % of its whole area under the low vulnerable zone. The contributing features in Physiographic vulnerability for the study area derived from the model are shown in Figure 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 Figure 12. 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 flooded 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 physiographic 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 physiographic 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 physiographic vulnerability has more importance since the area of impact will be based on the physiography. A combination of socio-economic and physiographic 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 physiographic and socio-economic vulnerabilities and the overall vulnerability to flood for the study area were 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 Figure 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 physiographic aspects of vulnerability, as illustrated in the Figure.

## Conclusion

The focus of the study lies in the assessment of vulnerability to floods by combining the two major criteria of physiographic 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 geo-spatial approach was made use of in the preparation of vulnerability maps. The contribution of various factors associated with socio-economic and physiographic 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 physiographical 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 physiographic vulnerable can be reduced.

## Abbreviations

GIS: Geographic Information System

MCDA: Multi-Criteria Decision Analysis

SV: Socio-economic Vulnerability

PV: Physiographic Vulnerability

NDVI: Normalised Differential Vegetation Index

LULC: Land use and Landcover

SPI: Stream Power Index

TWI: Topographic Wetness Index

CTI: Compound Topographic Index

DEM; Digital Elevation Model

AHP: Analytical Hierarchy Process

WLC: Weighted Linear Combination

CR: Consistency Ratio

## **Declarations**

### **Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### **Competing interests**

The authors declare that they have no competing interests

### **Funding**

Not applicable

### **Authors' contributions**

Not applicable

### **Acknowledgements**

The authors are thankful for the support from Aluva municipality office, Census of India, National Remote Sensing Centre (NRSC) and Kerala State Disaster Management Authority (KSDMA) for providing the necessary data required for the study

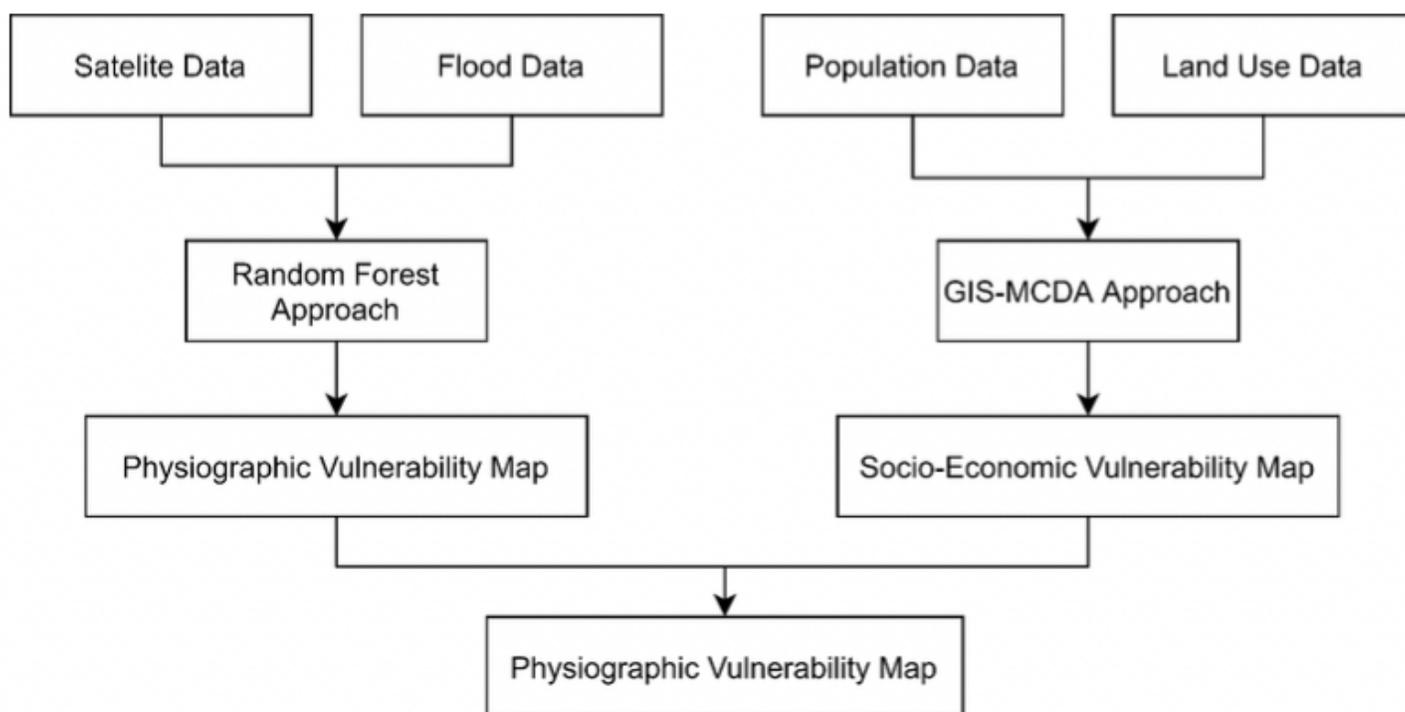
## References

1. Ai, F.F., Bin, J., Zhang, Z.M., Huang, J.H., Wang, J.B., Liang, Y.Z., Yu, L. and Yang, Z.Y., (2014). Application of random forests to select premium quality vegetable oils by their fatty acid composition. *Food chemistry*, 143, pp.472-478. doi:<https://doi.org/10.1016/j.foodchem.2013.08.013>
2. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32. doi:<https://doi.org/10.1023/A:1010933404324>
3. Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, 99(6), 323-329. doi:<https://doi.org/10.1016/j.ygeno.2012.04.003>
4. Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. doi:<https://doi.org/10.1890/07-0539.1>
5. De Brito, M. M., & Evers, M. (2016). Multi-criteria decision making for flood risk management: a survey of the current state-of-the-art. *Natural Hazards and Earth System Sciences Discussions*, 3, 6689-6726. doi: 5194/nhessd-3-6689-2015
6. 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 Res Manag* 21(9):1601–1612. doi: <https://doi.org/10.1007/s11269-006-9116-1>
7. Dong, L. J., Li, X. B., & Kang, P. E. N. G. (2013). Prediction of rockburst classification using Random Forest. *Transactions of Nonferrous Metals Society of China*, 23(2), 472-477. doi:[https://doi.org/10.1016/S1003-6326\(13\)62487-5](https://doi.org/10.1016/S1003-6326(13)62487-5)
8. Feizizadeh Bakhtiar & Stefan Kienberger(2017) Spatially explicit sensitivity and uncertainty analysis for multicriteria-based vulnerability assessment, *Journal of Environmental Planning and Management*, 60:11, 2013-2035. doi: <https://doi.org/10.1080/09640568.2016.1269643>
9. Fekete, A. (2009). Validation of a social vulnerability index in context to river-floods in Germany. *Natural Hazards & Earth System Sciences*, 9(2). doi:<https://doi.org/10.5194/nhess-9-393-2009>
10. 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. *Proceedings of the 7th Intl. Congress on Env. Modelling and Software*, San Diego, CA, USA. doi: <https://ssrn.com/abstract=2455821>
11. 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. *The Egyptian Journal of Remote Sensing and Space Science*, 15(2), 135-141. doi:<https://doi.org/10.1016/j.ejrs.2012.07.002>
12. Khosravi, K., Nohani, E., Maroufinia, E., & Pourghasemi, H. R. (2016). A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural Hazards*, 83(2), 947-987. doi:<https://doi.org/10.1007/s11069-016-2357-2>

13. Kirby, R. H., Reams, M. A., Lam, N. S., Zou, L., Dekker, G. G., & Fundter, D. Q. P. (2019). Assessing social vulnerability to flood hazards in the Dutch Province of Zeeland. *International Journal of Disaster Risk Science*, 10(2), 233-243. doi:<https://doi.org/10.1007/s13753-019-0222-0>
14. Lee, M. J., Kang, J. E., & Jeon, S. (2012). Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In 2012 IEEE international geoscience and remote sensing symposium (pp. 895-898). doi:1109/IGARSS.2012.6351414
15. Lee, M.J., Jung Eun Kang & Geunhan Kim (2015) Application of fuzzy combination operators to flood vulnerability assessments in Seoul, Korea, *Geocarto International*, 30:9, 1052-1075. doi:10.1080/10106049.2015.1027290
16. Lee, S., Kim, J. C., Jung, H. S., Lee, M. J., & Lee, S. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, 8(2), 1185-1203. doi:<https://doi.org/10.1080/19475705.2017.1308971>
17. Malczewski, J. (2006). Ordered weighted averaging with fuzzy quantifiers: GIS-based multicriteria evaluation for land-use suitability analysis. *International journal of applied earth observation and geoinformation*, 8(4), 270-277. <https://doi.org/10.1016/j.jag.2006.01.003>
18. Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrological processes*, 5(1), 3-30. doi:<https://doi.org/10.1002/hyp.3360050103>
19. Parker, D. J. (2000). Managing flood hazards and disasters: international lessons, directions and future challenges. *Floods*, 2, 287-306.
20. Paulo Fernandez, Sandra Mourato and Madalena Moreira (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 and Risk* doi:10.1080/19475705.2015.1052021
21. Pourghasemi, H. R., Jirandeh, A. G., Pradhan, B., Xu, C., & Gokceoglu, C. (2013). Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2), 349-369. doi: <https://doi.org/10.1007/s12040-013-0282-2>
22. Saaty T (1977) " A scaling method for priorities in hierarchical structures." *Journal on Mathematical Psychology*. 15(3):234\_281, [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5)
23. Saaty T. 1980. *The analytic hierarchy process: planning, priority setting, resource allocation*. New York, NY: McGraw-Hill.
24. Samanta, S., Pal, D. K., & Palsamanta, B. (2018). Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Applied Water Science*, 8(2), 66. doi:<https://doi.org/10.1007/s13201-018-0710-1>
25. Sarkar, D., & Mondal, P. (2020). Flood vulnerability mapping using frequency ratio (FR) model: a case study on Kulik river basin, Indo-Bangladesh Barind region. *Applied Water Science*, 10(1), 17. doi:<https://doi.org/10.1007/s13201-019-1102-x>
26. Solomon, T., & Liu, Z. (2010). Earthquake induced damage classification for reinforced concrete buildings. *Structural safety*, 32(2), 154-164. doi:<https://doi.org/10.1016/j.strusafe.2009.10.002>

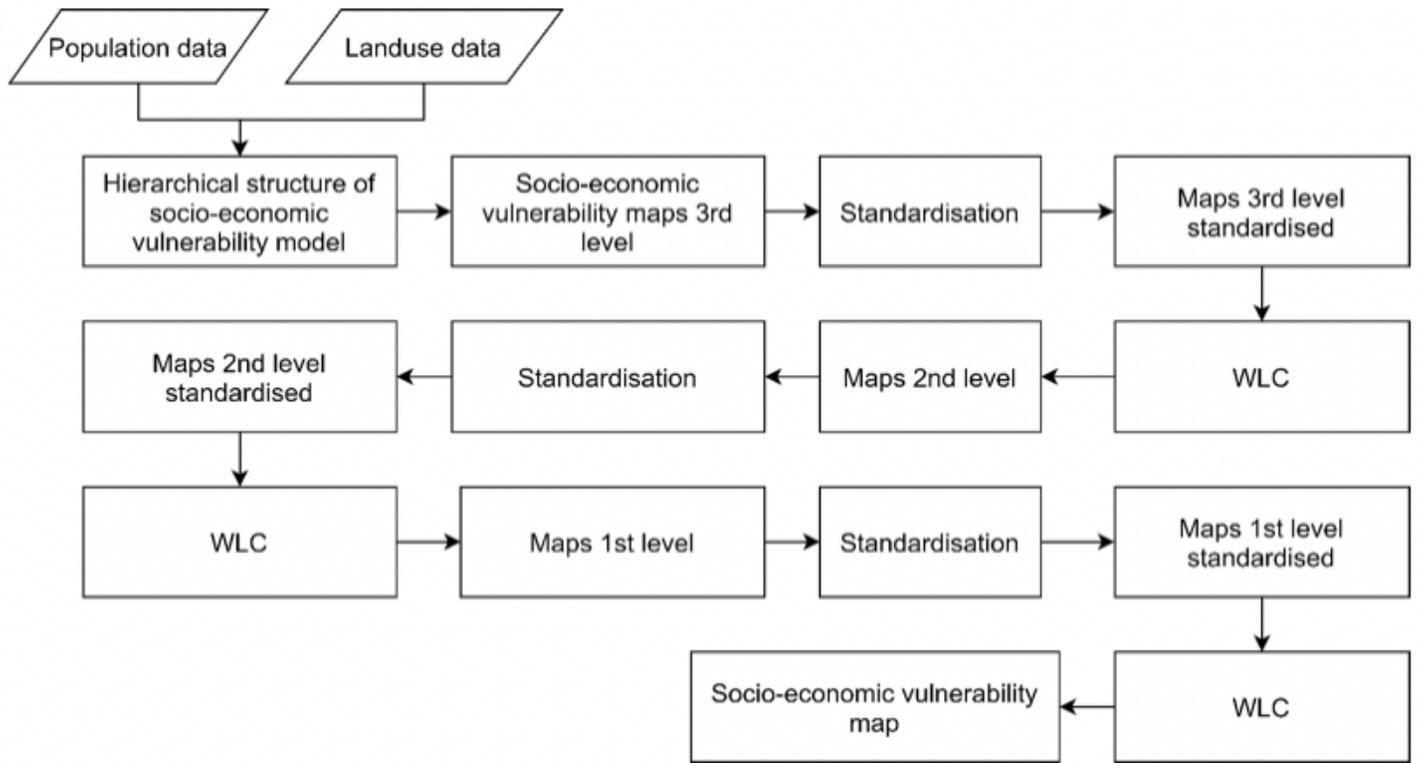
27. Stefanidis, S., & Stathis, D. (2013). Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Natural hazards*, 68(2), 569-585. <https://doi.org/10.1007/s11069-013-0639-5>
28. Tsoutsos, T., Drandaki, M., Frantzeskaki, N., Iosifidis, E., & Kiosses, I. (2009). Sustainable energy planning by using multi-criteria analysis application in the island of Crete. *Energy policy*, 37(5), 1587-1600. doi:<https://doi.org/10.1016/j.enpol.2008.12.011>
29. Wang, Q., Watanabe, M., Hayashi, S., & Murakami, S. (2003). Using NOAA AVHRR data to assess flood damage in China. *Environmental monitoring and assessment*, 82(2), 119-148. doi:<https://doi.org/10.1023/A:1021898531229>
30. Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130-1141. doi:<https://doi.org/10.1016/j.jhydrol.2015.06.008>
31. Yeh, C. C., Chi, D. J., & Lin, Y. R. (2014). Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, 254, 98-110. doi:<https://doi.org/10.1016/j.ins.2013.07.011>

## Figures



**Figure 1**

Methodology for Vulnerability



**Figure 2**

Methodology for Socio-economic vulnerability

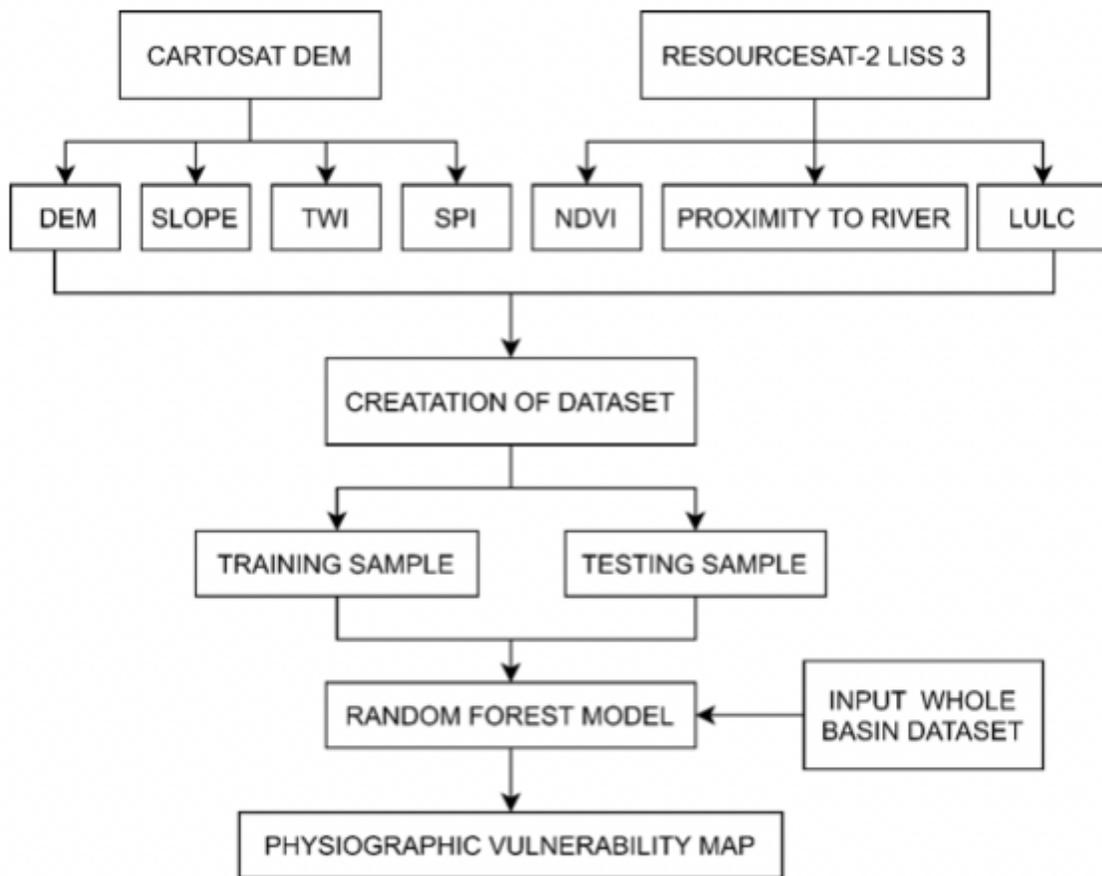


Figure 3

Methodology for physiographic vulnerability

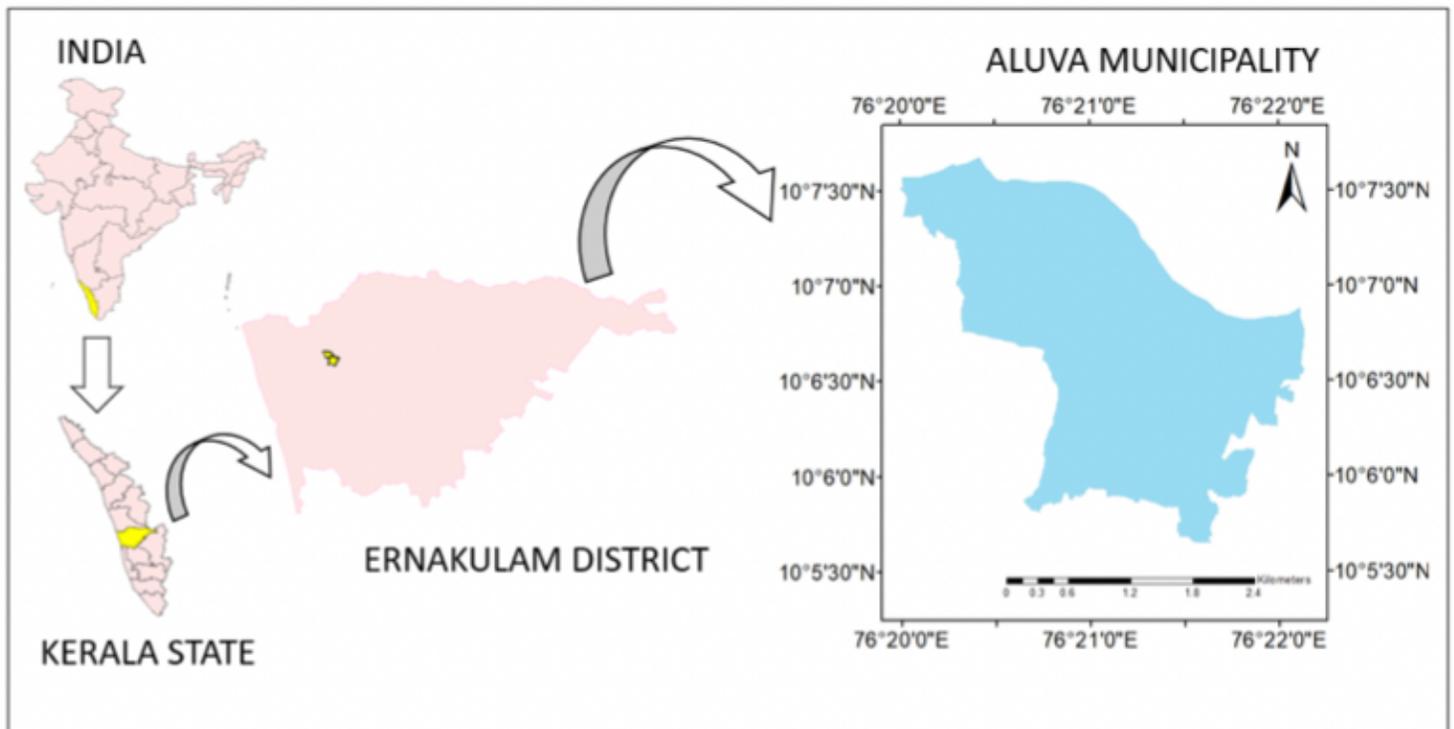


Figure 4

Study area

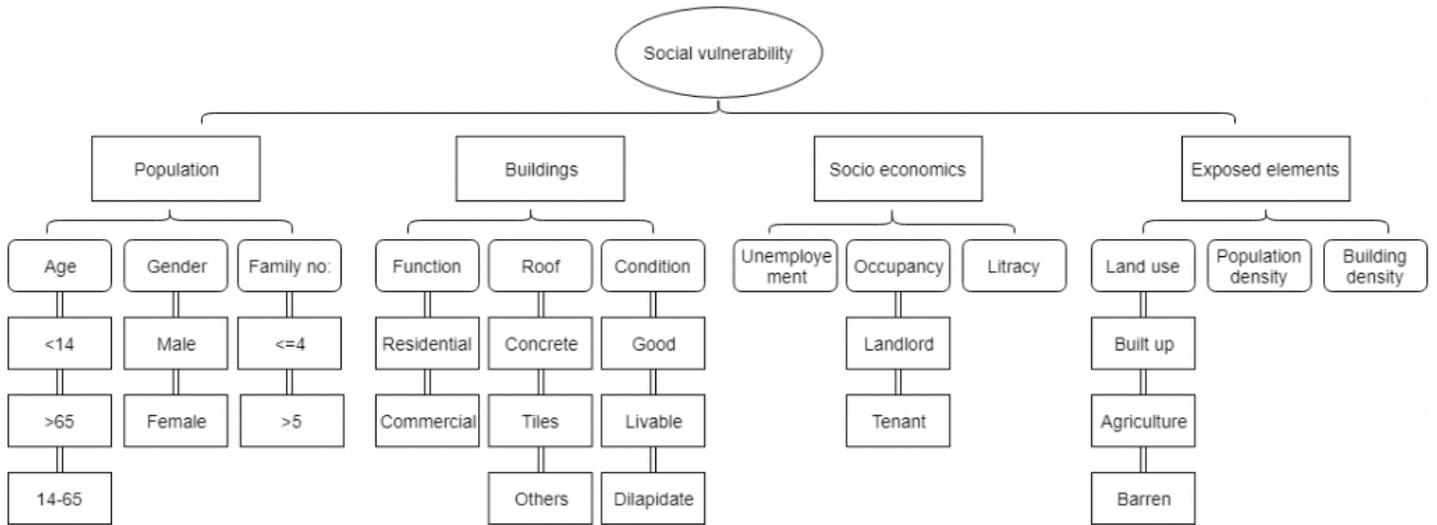
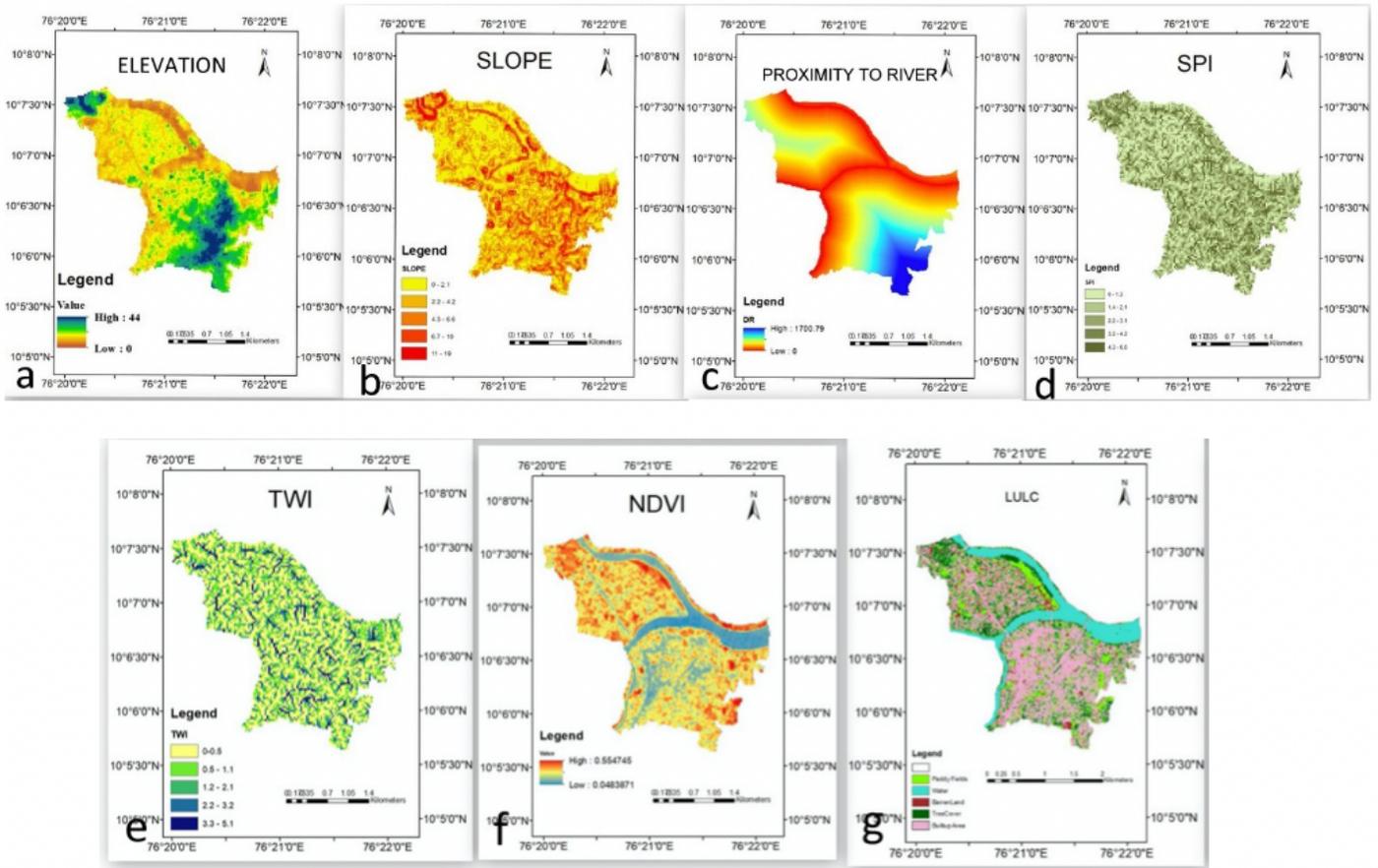


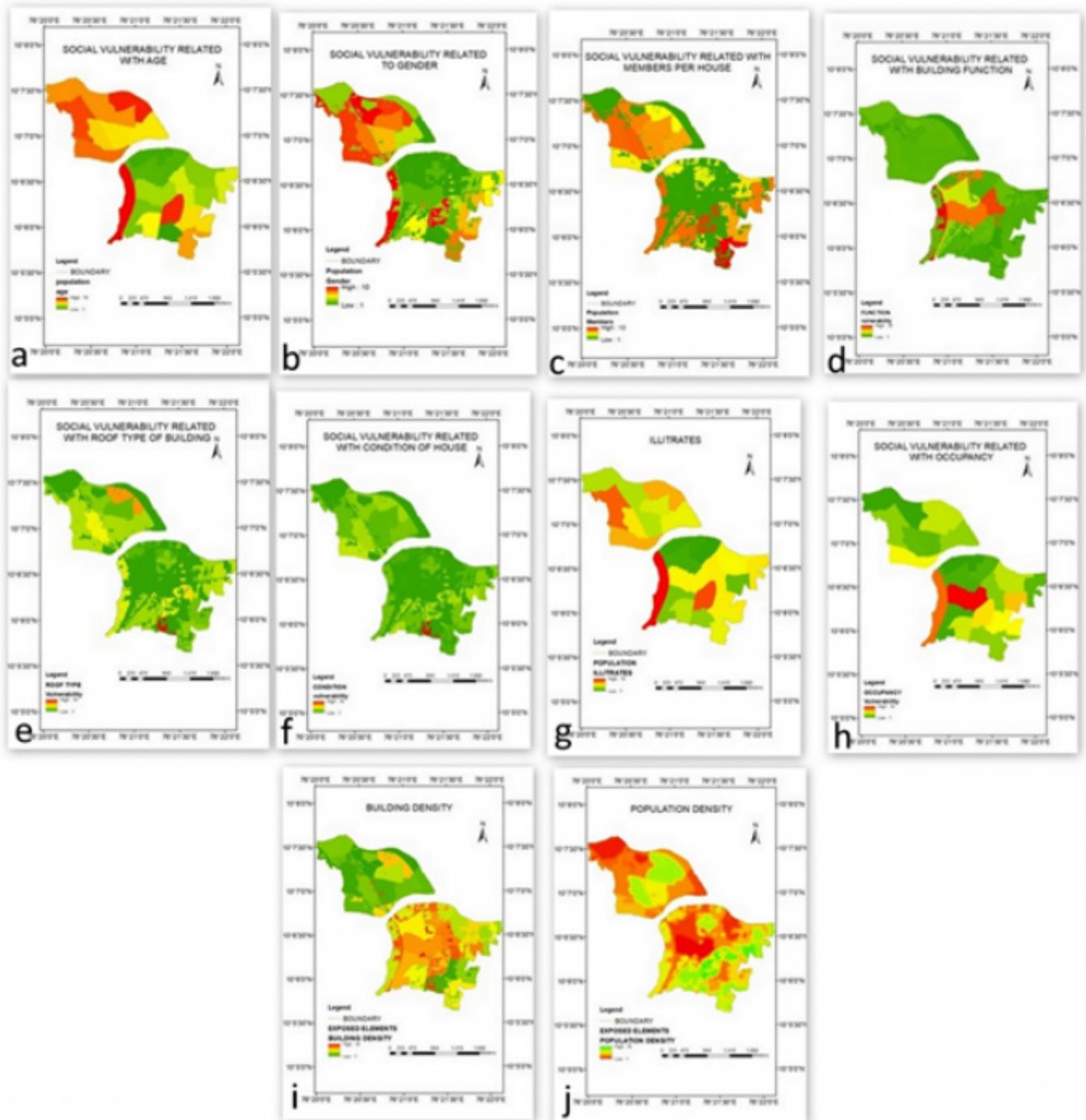
Figure 5

Hierarchical structure of socio-economic vulnerability model



**Figure 6**

Raster maps created for factors: (a) Elevation (b) slope (c) proximity to river, (d) SPI (e) TWI (f) NDVI (g) LULC



**Figure 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

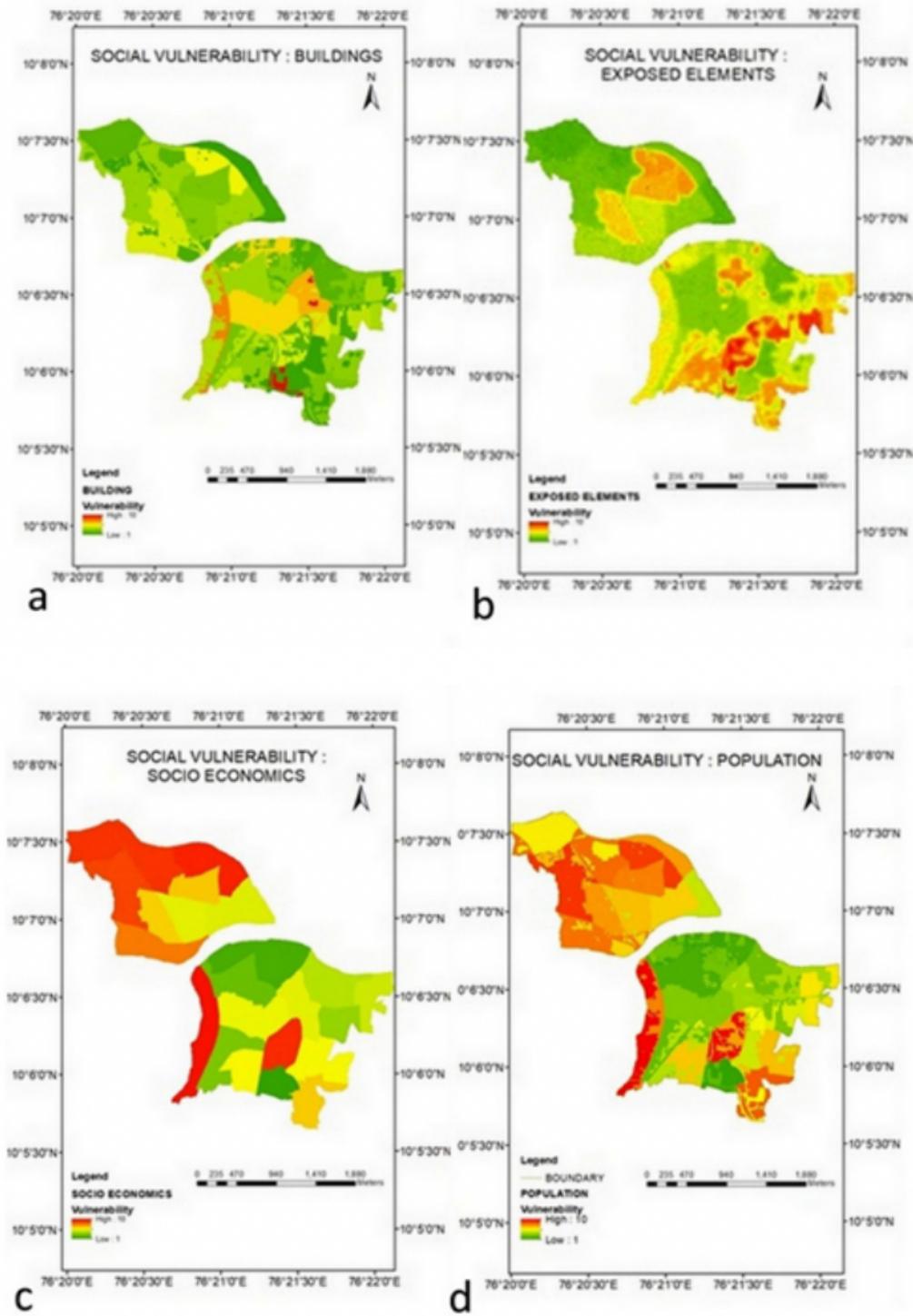
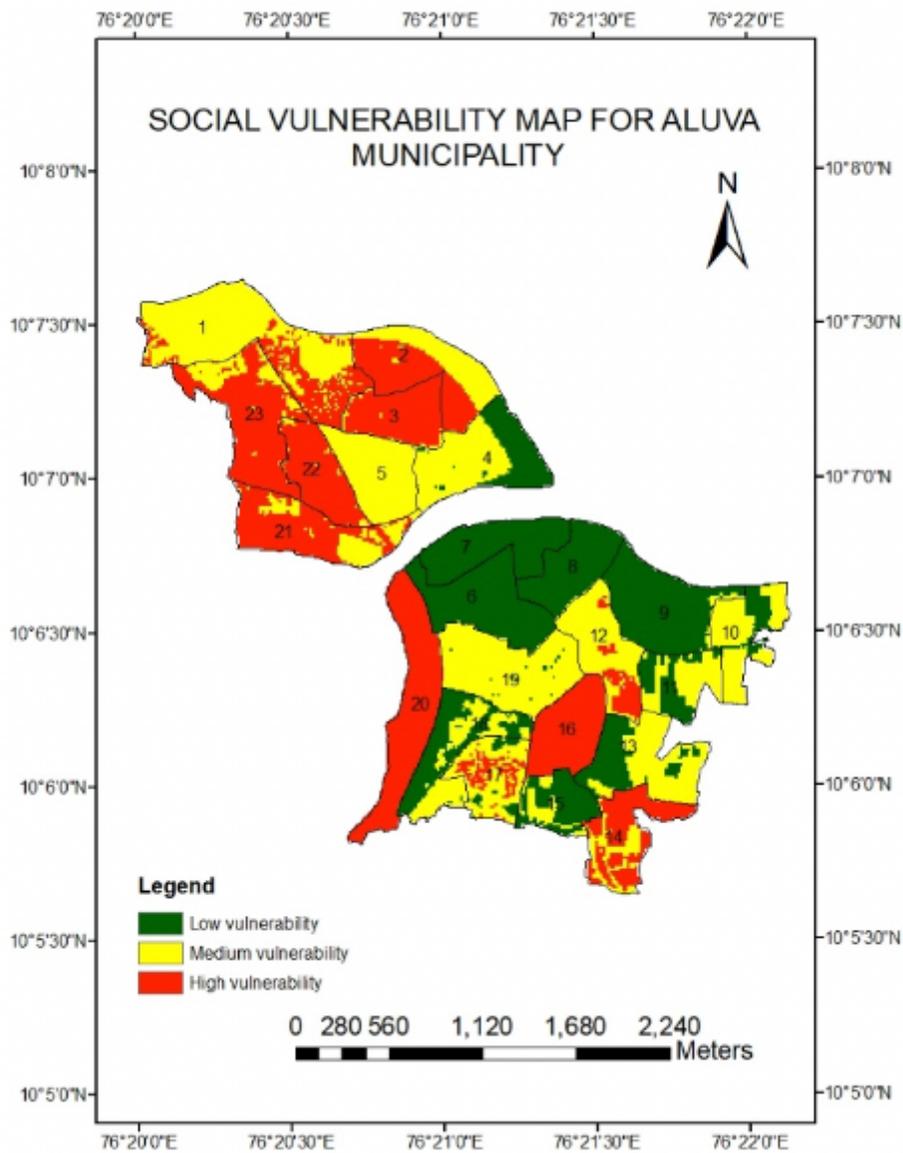


Figure 8

Socio-economic vulnerability maps showing four objectives (a) Building (b) Exposed elements (c) Socioeconomics (d) Population



**Figure 9**

Socio-economic Vulnerability map of for Aluva municipality

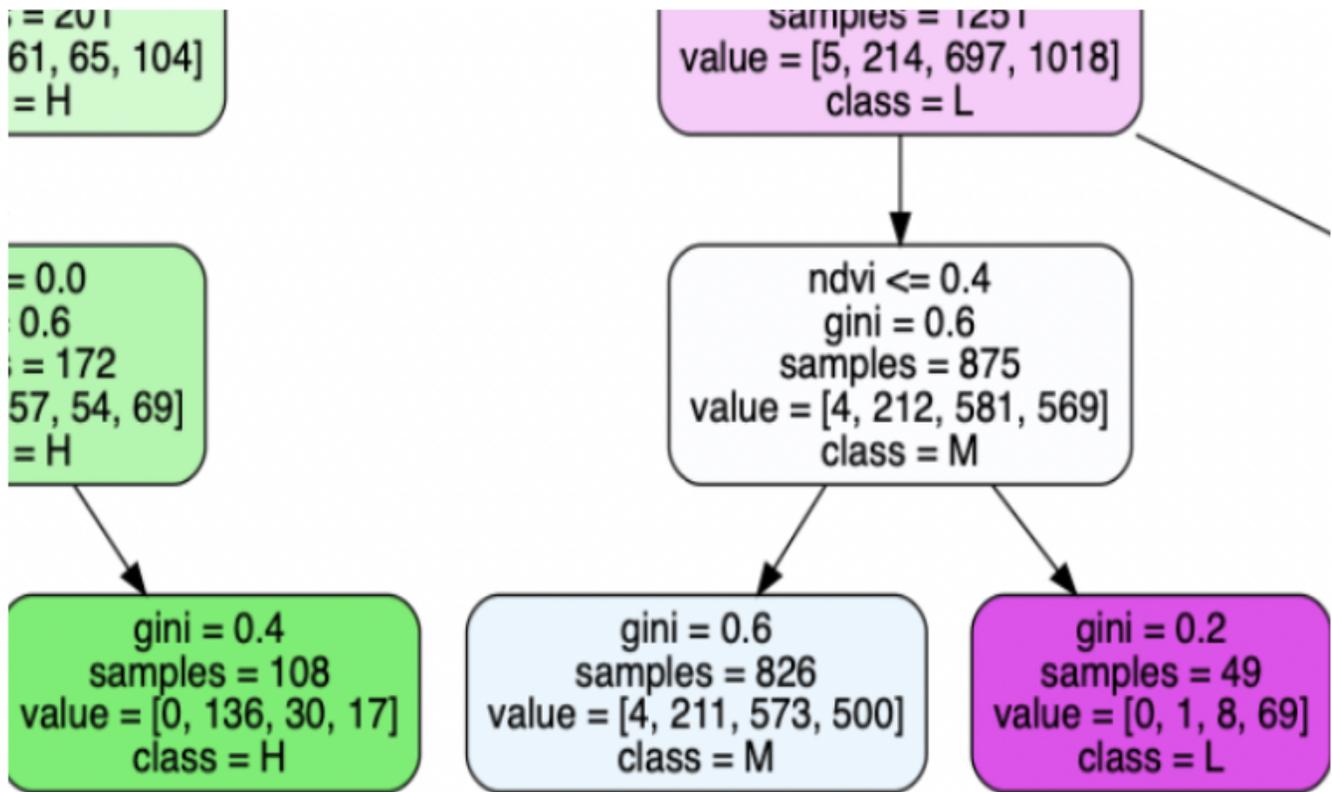


Figure 10

Visualization of the Decision tree (leaf nodes)

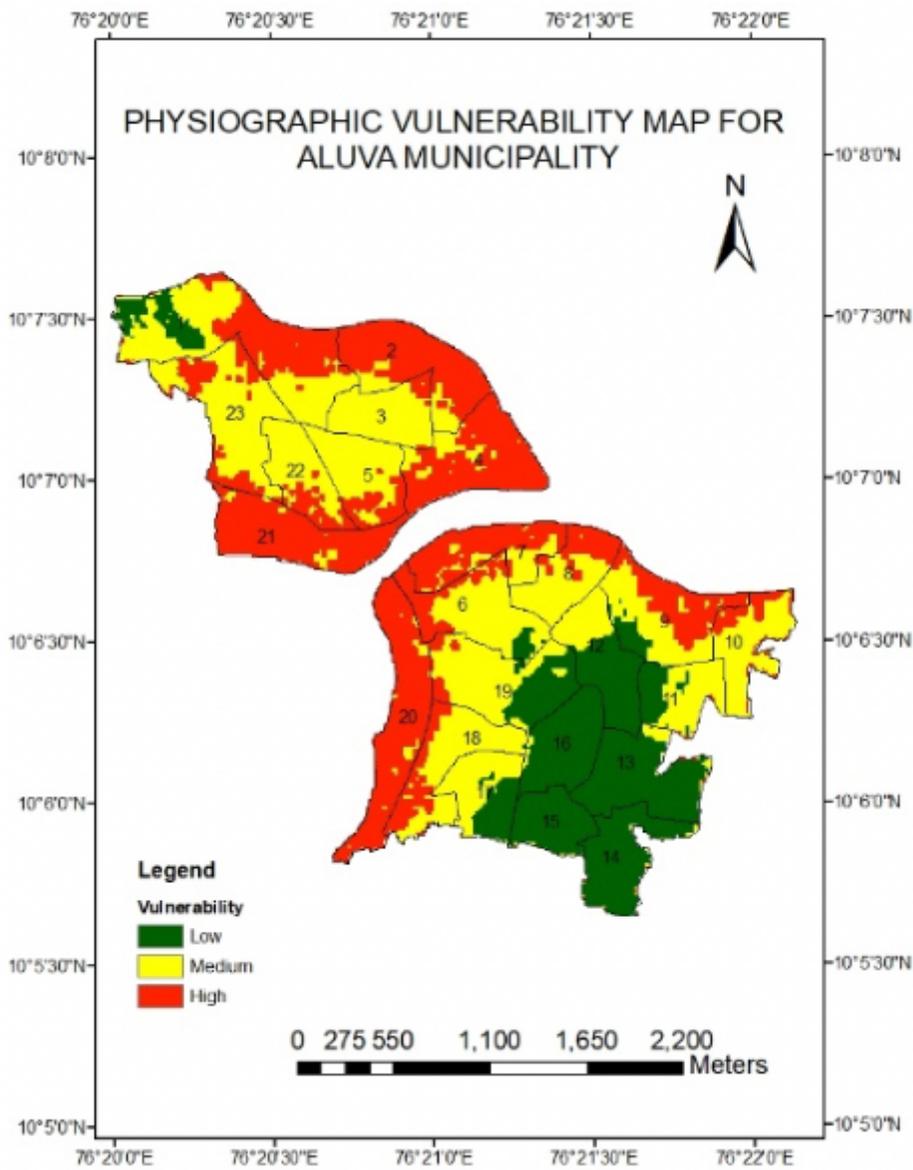


Figure 11

PV for Aluva municipality

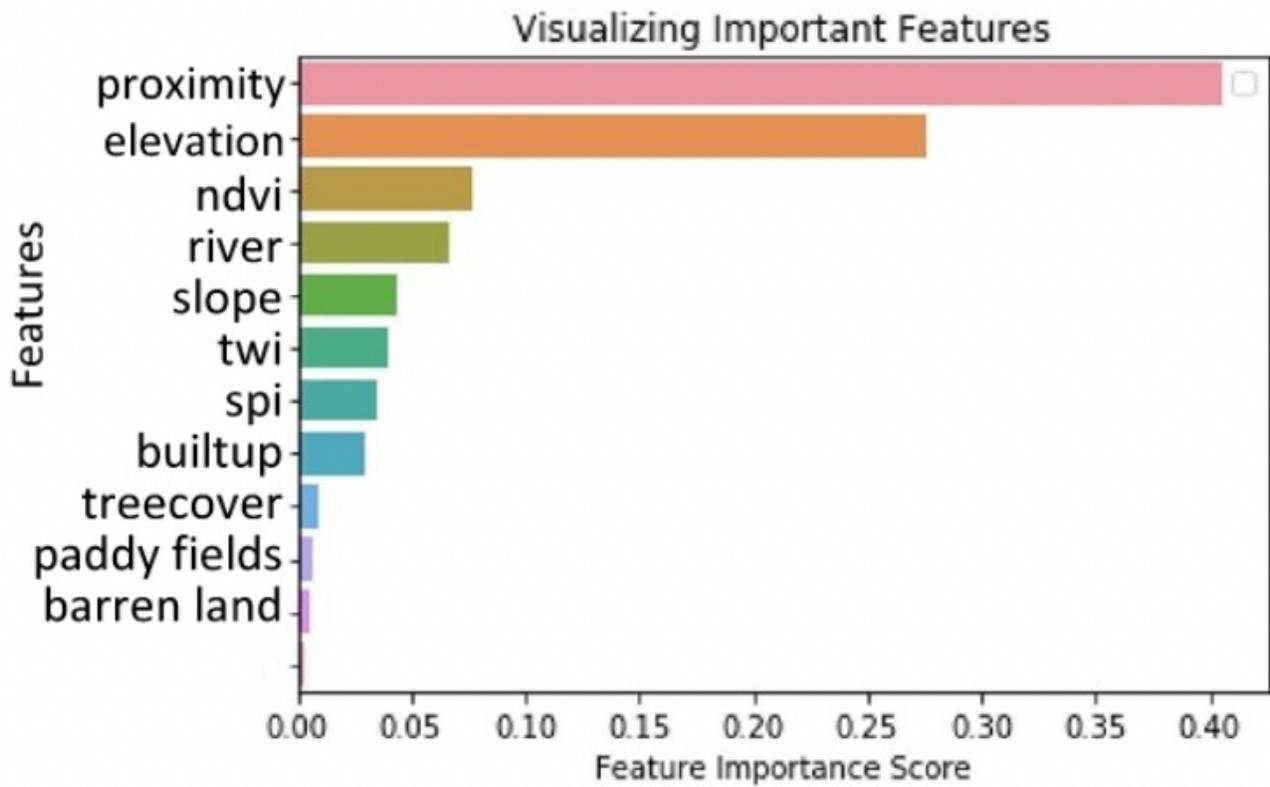
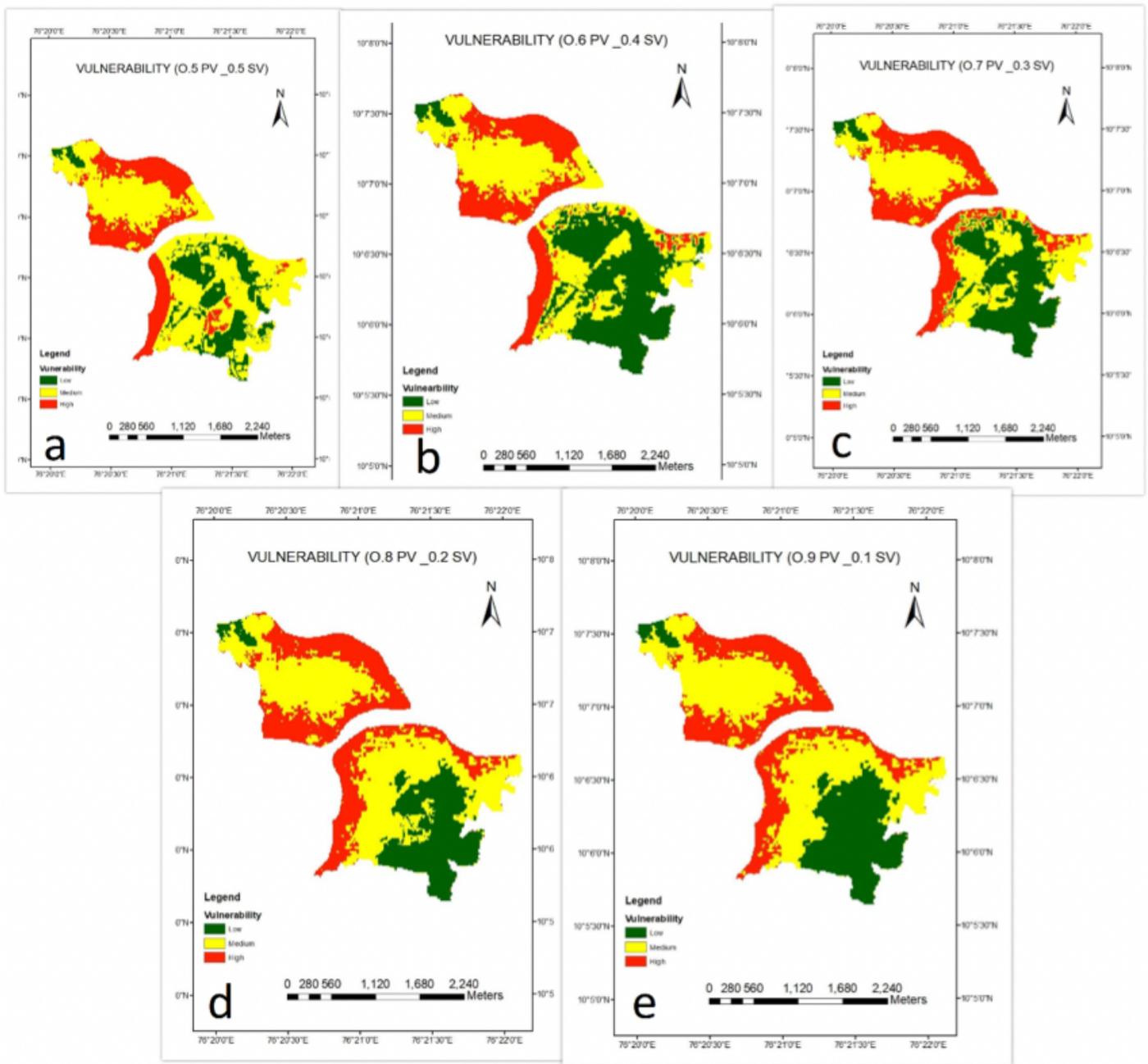


Figure 12

Indices importance degree and contribution of indices



**Figure 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