Landslide Risk Zoning Applying Kohonen's Self-Organizing Map Neural Network Technique

Bayes Ahmed[1]

Renato Forte[2]

Abstract

Every year during the monsoon period, the vulnerable communities living in the dangerous hill slopes in Cox's Bazar Municipality (CBM), Bangladesh face landslide hazards. The frequency and intensity of landslide hazards are increasing due to torrential rainfall in short period of time. In addition, being the most attractive tourist city of Bangladesh, CBM is facing acute population pressure. The rate of urbanization is also high in CBM. The local people are building residential houses by cutting the hills and making the landslide disaster scenario worse. Frequent landslide hazards are causing human casualties, and property, infrastructure and roads are damaged. The economic loss due to landslides is also high in CBM. Therefore, it is necessary to prepare the landslide risk zoning maps to help mitigate the adverse impacts. This article has adopted the Kohonen's Self-Organizing Map (SOM) neural network technique. To produce the SOM for landslide risk mapping, a total of 12 factor maps (i.e. slope, land cover, geology, geomorphology, NDVI, soil moisture, rainfall pattern, and distance from-existing buildings; stream, road and drainage network, and faults and lineaments) are selected. A detail landslide inventory map is prepared for network training and model validation purpose. The performance of the SOM method is validated using the Area Under the relative operating characteristic Curve (AUC) method. The AUC value is calculated 86.60%. The SOM image is reclassified; applying natural breaks (Jenks) method, in four different risk zones - very high, high, medium and low. The landslide risk-zoning map as produced by SOM technique is found scientifically significant and representing the ideal situation. The concerned urban planners, engineers and stakeholders can use this kind of risk-zoning map for formulating policies related to landslide disaster risk reduction in the hill districts of Chittagong, Bangladesh.

Key words: landslide risk zoning, self-organizing map, neural network, ROC curve.

1. Introduction

Landslides are common hazards in the south-eastern Chittagong Hill Districts (CHD) of Bangladesh (Figure 1a). The landslides mainly occur during the monsoon (June-August) with torrential rainfall in short time span (2-4 days). Although CHD is located in high-risk zone for earthquakes but still no evidence is found on earthquake-triggered landslides in the region. Among the five hill districts of Chittagong, Cox's Bazar Municipality (CBM) is highly vulnerable to landslide disasters (Ahmed, 2015a). For example, rainfall triggered landslides caused at least 47 casualties in CBM on 15 June 2010. Moreover, there were about 15 casualties, and extensive level of infrastructure and property damaged in Cox's Bazar district in the months of June and July 2015 (bdnews24, 2015) due to flash flood, storm and landslides. It is evident that CBM is facing rainfall associated landslide disasters every year. Moreover, extensive hill cutting for developing residential houses (Figure 2), institutional weaknesses and population pressure are exaggerating the landslide disaster scenario in CBM. It is, therefore, necessary to prepare a scientifically valid landslide hazard-zoning map for CBM to help reducing the negative impacts landslide disasters.

[1]PhD Candidate, Institute for Risk and Disaster Reduction, University College London (UCL), London, UK. Email: bayesahmed@gmail.com; bayes.ahmed.13@ucl.ac.uk

[2]GIS Consultant, Cyient Europe Ltd., BT Telephone Exchange, New Street, HP20 2NN, Aylebsury, UK. Email: renato_forte@virgilio.it

CBM is bounded on the west by the Bay of Bengal, on the northeast by Bakkhali River, and on the north by Moheshkhali Channel (Figure 1b). The total area of CBM is about 20.78 sq.km. The population of CBM increased fourfold in the past two decades (1991–2011), which is now around 1,67,477 (Ahmed, 2015a). The primary objective of this paper is to prepare a landslide risk-zoning map for CBM using the Kohonen's self-organizing map neural network technique.

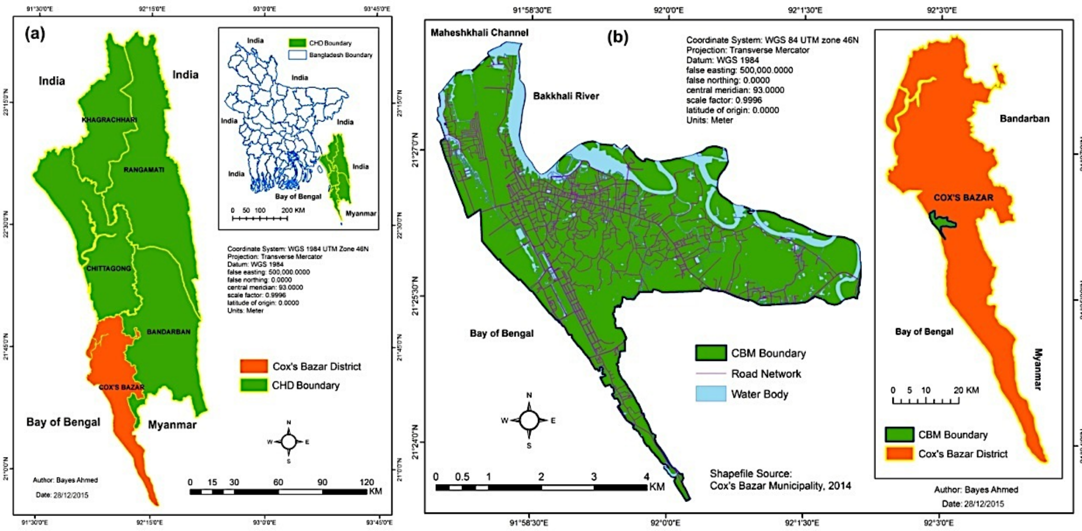

Figure 1. (a) Location of Chittagong hill districts in Bangladesh; **(b)** Location of Cox's Bazar Municipality in Cox's Bazar district. Source: CBM and Bayes Ahmed, 2015.

2. Literature Review and Data Collection

Landslide is the movement of mass of soil or earth, down a slope, when shear stress exceeds the shear strength of the material (van Westen et al., 2011, pp. 3-55). There are many popular techniques in Landslide Susceptibility Mapping (LSM) that can be later converted into landslide risk zoning maps for a particular area. In recent times, the use of Geographic Information System (GIS), Remote Sensing (RS) and spatial statistical techniques are being popular in LSM. Some techniques are quantitative data driven and some are qualitative weight based techniques (Guzzetti et al., 1999). For example: multi-criteria decision analysis methods—the Artificial Hierarchy Process (AHP), Weighted Linear Combination (WLC), and Ordered Weighted Average (Ahmed, 2015b). Binomial logistic regression, multinomial logistic regression, multiple linear regression and Multi-Layer Perceptron (MLP) neural network classifier using the back propagation algorithm are some of the examples of data driven statistical techniques for LSM (Eastman, 2012; Park et al., 2013).

For this research purpose, 12 landslide causative factor maps and a landslide inventory map was produced. All the images were projected to the universal transverse Mercator (UTM) zone 46 North system, with the world geodetic system (WGS)-1984 Datum. Each image resolution was set to cell size of 30 meter, with 268 columns and 290 rows. The digital elevation model (DEM) image, dated on 29 November 2013, was acquired from the advanced space-borne thermal emission and reflection radiometer (ASTER) global digital elevation model web-portal. The same ASTER image was used for generating slope (Figure 3a) and stream network (Figure 3e) layers. The land cover map (Figure 3c) with four broad land cover classes (i.e. builtup area, vegetation, agricultural and fallow land, and water body) was prepared using the Landsat 8 Operational Land Imager (OLI) image (dated 24 April 2014). The same OLI image was used for generating the Normalized difference vegetation index (NDVI) layer (Figure 3d).

Figure 2. Photographs showing location of residential houses in dangerous hill slopes in Lighthouse Para, Cox's Bazar Municipality, Bangladesh. Source: Bayes Ahmed, 2014.

The precipitation map of CBM (Figure 3b) is prepared using the collected daily precipitation data (1950–2010) from the Bangladesh Meteorological Department. The road (Figure 3g), drainage network (Figure 3h), and existing structure (Figure 3f) layers were collected from CBM. The geological (Figure 3j), geomorphological (Figure 3i), soil moisture (Figure 3l), and fault–lineaments (Figure 3k) layers were collected from the Geological Survey of Bangladesh. Euclidean distance technique was implemented to generate the distance images. To prepare the landslide inventory map (Figure 4), extensive fieldwork was carried out in CBM from July to October 2014. A total of 74 landslides were identified in CBM (Figure 5). A handheld global positioning system (GPS) device was used to identify the locations of the landslide hazards.

3. Methodology

This methodology section is written from the IDRISI Selva help system (Eastman, 2012). The tool used for landslide hazard zoning is known as Self-Organizing Map (SOM). SOM undertakes both unsupervised and supervised classification of remotely sensed imagery using Kohonen's Self-Organizing Map (SOM) neural network. The SOM procedure used here is closely modelled after the Kohonen (1990) procedure. Figure 5 illustrates the basic architecture of a SOM. The input layer represents the input feature vector and thus has neurons for each measurement dimension. In the case of remotely sensed data, this would imply a separate neuron for each reflectance band. The output layer of a SOM is typically organized as a two-dimensional (typically square) array of neurons, although one-dimensional structures are common. Each output layer neuron is connected to all neurons in the input layer by synaptic weights.

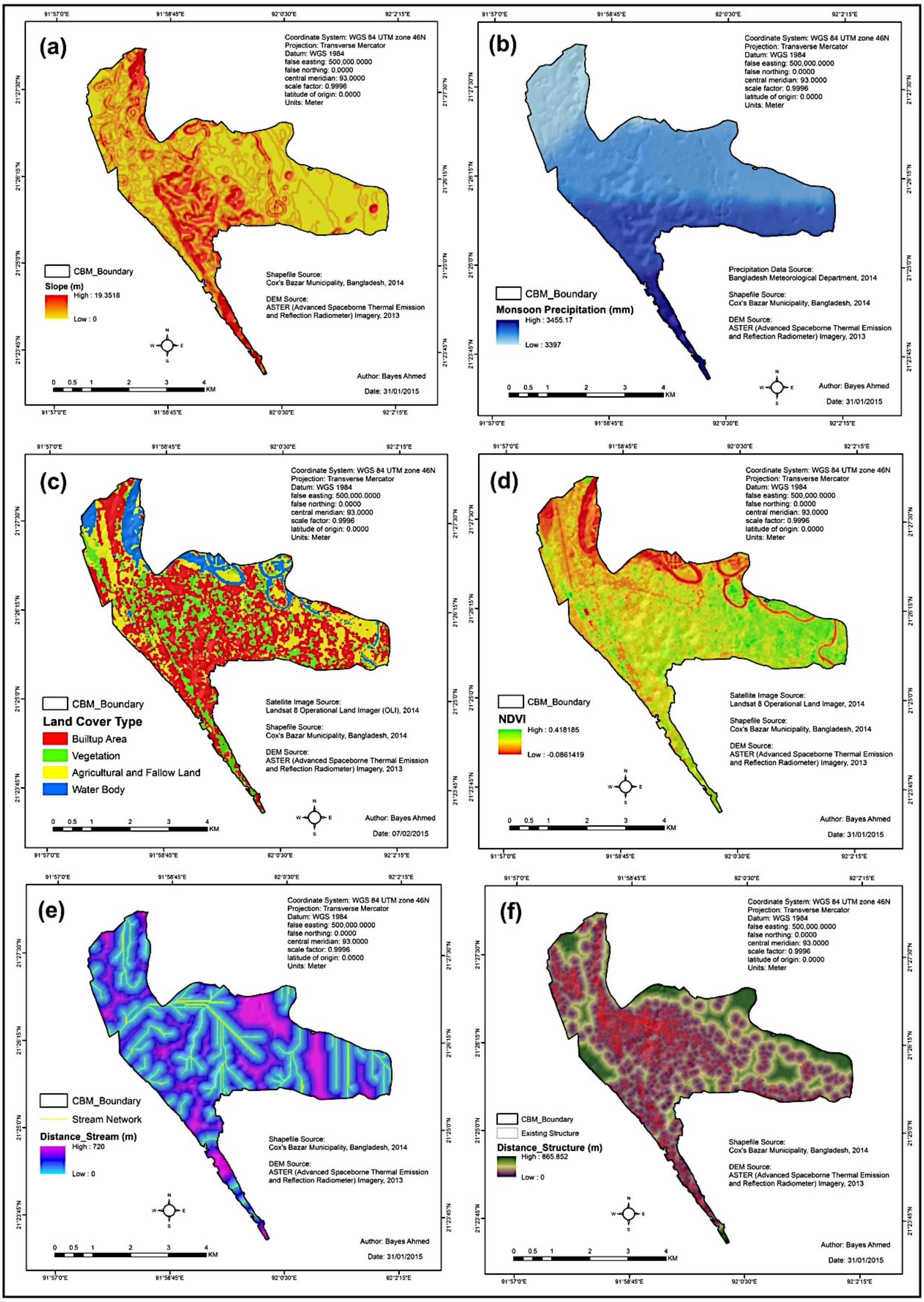

Figure 3. (a) Slope, **(b)** rainfall pattern, **(c)** land cover, **(d)** NDVI, **(e)** distance from stream, and **(f)** distance from existing building structure maps of CBM. Source: Ahmed, 2015a.

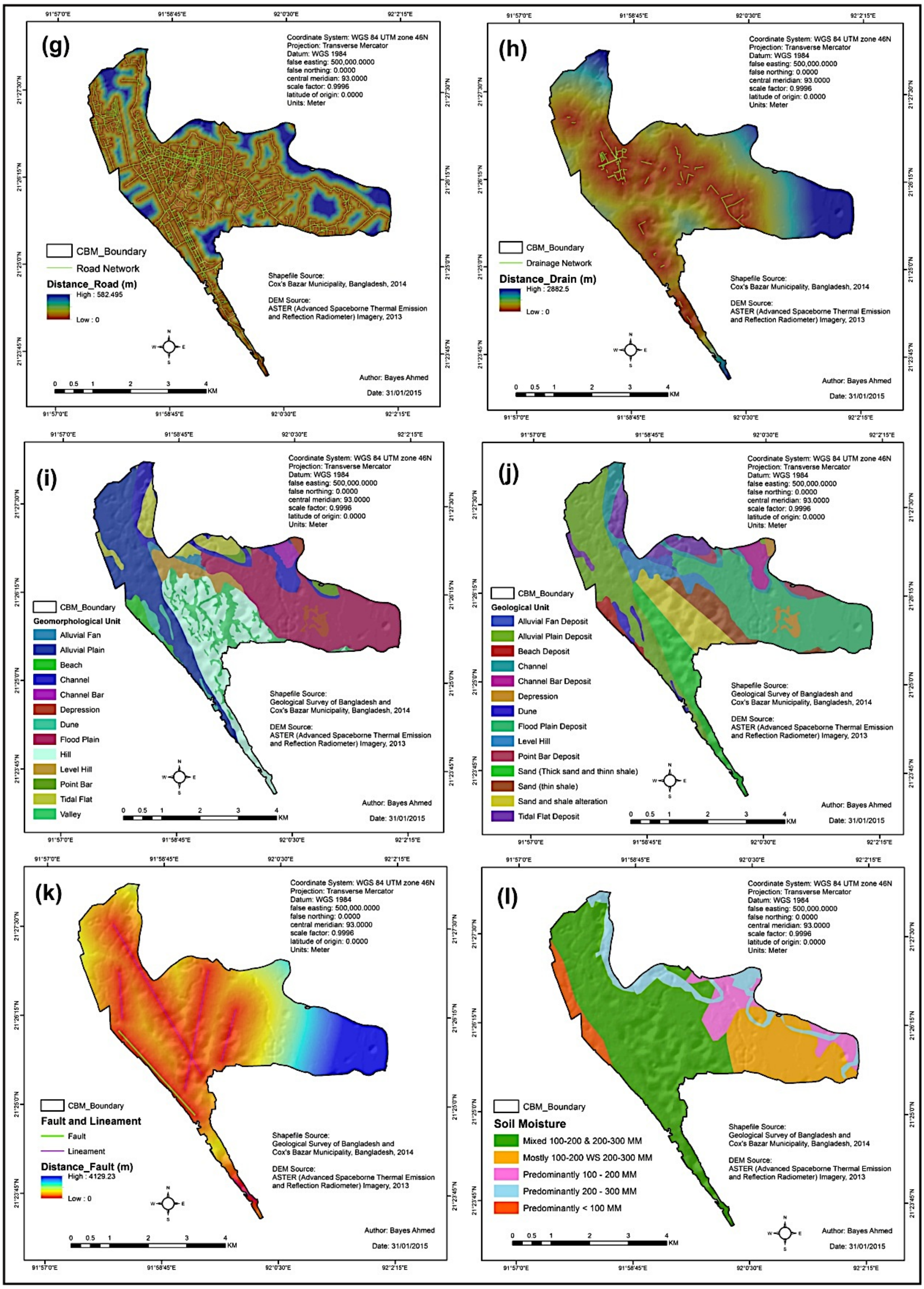

Figure 3. (g) Distance from road network, **(h)** distance from drainage network, **(i)** geomorphology, **(j)** geology, **(k)** distance from faults and lineaments, and **(l)** soil moisture maps of CBM. Source: Ahmed, 2015a.

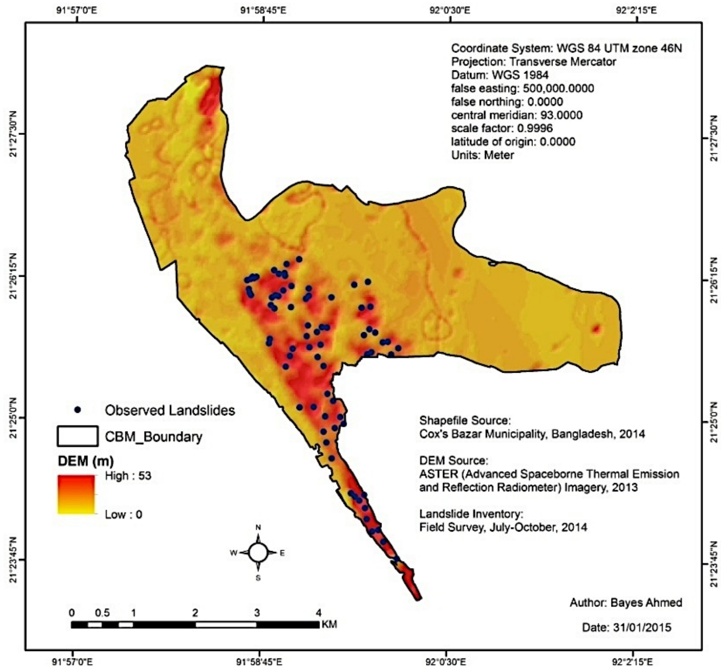

Figure 4. Landslide inventory map of CBM. Source: Ahmed, 2015a.

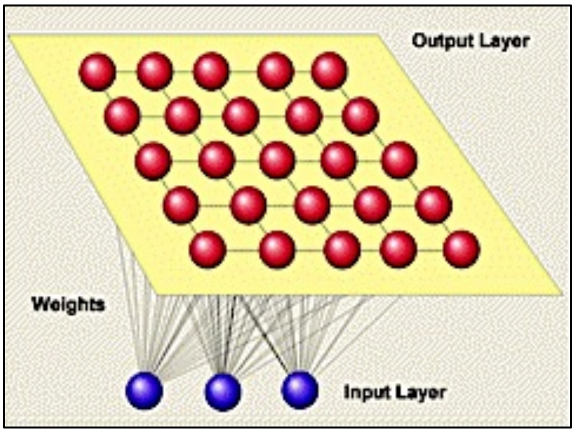

Figure 5. Example of the architecture of a SOM with an input layer (made up of three neurons) and an output layer (made up of 5 × 5 neurons). Source: Eastman, 2012.

In its use for supervised classification, the procedure begins with a coarse tuning phase that is effectively a form of unsupervised classification. A later fine tuning stage refines intra-class decision boundaries using a Learning Vector Quantization (LVQ) procedure. Another critical phase between the coarse tuning and fine-tuning stages is the assignment of neurons (sub-clusters) to training classes – a process known as *labelling*.

Coarse tuning is an unsupervised classification stage in which competitive learning and lateral interaction lead to a fundamental regional organization (a topology) of neuron weights that represent the underlying clusters and sub-clusters in the input data. Before one can do this, it is necessary to set the input data and the synaptic weights with the same measurement range and scale. In our case, we chose to rescale the image digital numbers (DN's) to match the 0 - 1 range for the synaptic weights. Initially, synaptic weights between the output and input neurons are randomly assigned.

Before an SOM can perform a classification, code books must be labelled. This is to determine the class to which each output neuron belongs. At this stage, the basic structure of the input data has been topologically organized within the SOM output layer. The result, however, is similar to a cluster analysis in that the identity of regional groupings of neurons is unknown. The labelling stage is intended to establish the identities of these regional associations by comparison with training data. To do this, training data are fed to the coarse tuned network. The training site class that is assigned most frequently to a neuron then becomes its label. This procedure is known as the majority voting (Tso and Mather, 2001).

In some cases, one expects to improve accuracy by using a supervised training if there are training data available. The goal of fine tuning is to refine the decision boundaries between classes based on the training site data. In the coarse tuning stage, raw input data are fed into the training process. The topological organization of the feature map is thus based on global image characteristics and it is normal to find that a group of neurons is labelled with a single information class. That is, a group of neurons will commonly cover the range of variability in reflectance associated with the information class. In the fine tuning stage, the specific boundaries between neurons associated with specific information classes are refined using training site data.

In the procedure outlined by Kohonen (1990), fine-tuning is achieved through the use of Learning Vector Quantization (LVQ). With LVQ2, weight vectors are updated in the following way:

$$w_i^{t+1} = w_i^t - \delta^t (x_i - w_i^t) \qquad \text{and}$$

$$w_j^{t+1} = w_j^t + \delta^t (x_i - w_j^t) \qquad \text{if Ci is the nearest class, but x belongs to Cj (Cj} \neq \text{Ci)}$$

$$w_k^{t+1} = w_k^t \qquad \text{in all other cases}$$

The primary controls for the fine-tuning are thus the number of iterations and the gain factor δ^t setting.

With a converged status, SOM can characterize the distribution of input samples, and thus generate a two-dimensional map from a multi-dimensional feature space. The feature map of the SOM is analogous to the mapping of the cortex of the human brain upon which processing of spatial information is based. After either the coarse tuning and labelling or fine-tuning stages, the feature map can be viewed. The feature map is colour coded by information classes. Thus all neurons associated with a particular class share the same colour.

For supervised classification, this clustering is based on association with training site data. Therefore, for the classification of unlabelled neurons, an auxiliary algorithm is employed using the logic of agglomerative clustering to assign unlabelled neurons to the clusters already formed from the supervised stage. To establish a class for pixels that trigger the second form of disconnected units in the SOM feature map during classification, it was decided to evaluate the mean distance between the input pattern and all neurons associated with each class, i.e.,

$$\overline{D}_i = \frac{\sum_{j=1}^{N} \|x - w\|}{N_i} \qquad \forall Label(j) = i$$

Where x is the input vector, and w the organized reference vector (weight vector). N_i is the number of the output layer neurons that are labelled by class i (i = 1, 2, … m).

To determine the statistical reliability of the results, the area under the relative operating characteristic (ROC) curve (AUC) method was employed. The AUC is a good indicator to evaluate the performance of a model qualitatively, and it is being widely used for LSM validation. AUC values ≤ 0.5 indicate no improvement, between 0.7 and 0.9 indicate reasonable agreement, and AUC ≥ 0.9 represents ideal situation (Kavzoglu et al., 2014).

4. Results and Discussions

For training the network, the classification mode was specified as supervised classification. Then the 12 factor layers were added for SOM analysis. The network parameters were as follows:

Input Layer Neurons: 12, Output Layer Neurons: 225, Initial neighbourhood radius: 22.209999, Min Learning Rate: 0.500000, Max Learning Rate: 1.000000, Min Gain Term: 0.000100, and Max Gain Term: 0.000500.

Landslide inventory image was selected as the training site image. The fine-tuning rule was LVQ2 and fine tuning epochs were 50. For the soft classification map the minimum distance algorithm was chosen. The running statistics were found as- Learning rate: 0.50, Radius: 1.0, Gain term: 0.0005, Iteration: 50, and Quantization error: 0.2844. A screenshot of the SOM map preparation tool in IDRISI Selva software is depicted in Figure 6.

For model validation purpose, the SOM generated image was chosen as input image and the landslide inventory map (Figure 4) was selected as reference image. The AUC is found as 0.866. With each threshold, the following 2x2 contingency Table 1 is calculated:

Table 1. Reference image of threshold cells calculation.

Reality (reference image)		
Simulated by threshold	1	0
1	A (number of cells)	B (number of cells)
0	C	D
For the given reference image:	A+C=74	B+D=23020

Source: Bayes Ahmed, 2015.

Table 2 showing the values achieved for the true and false positive for ROC analysis.

Table 2. Calculation of true and false positive for ROC.

No.	Expected Thresholds (%)	Actual Thresholds (%)	Actual raw cuts	A	True Positive (%)	B	False Positive (%)
1	0	0	0	0	0	0	0
2	25	25	0.9707	71	95.9459	5703	24.7741
3	50	50	0.4824	74	100	11474	49.8436
4	75	75	0.2498	74	100	17247	74.9218
5	100	100	0.0108	74	100	23020	100

Source: Bayes Ahmed, 2015.

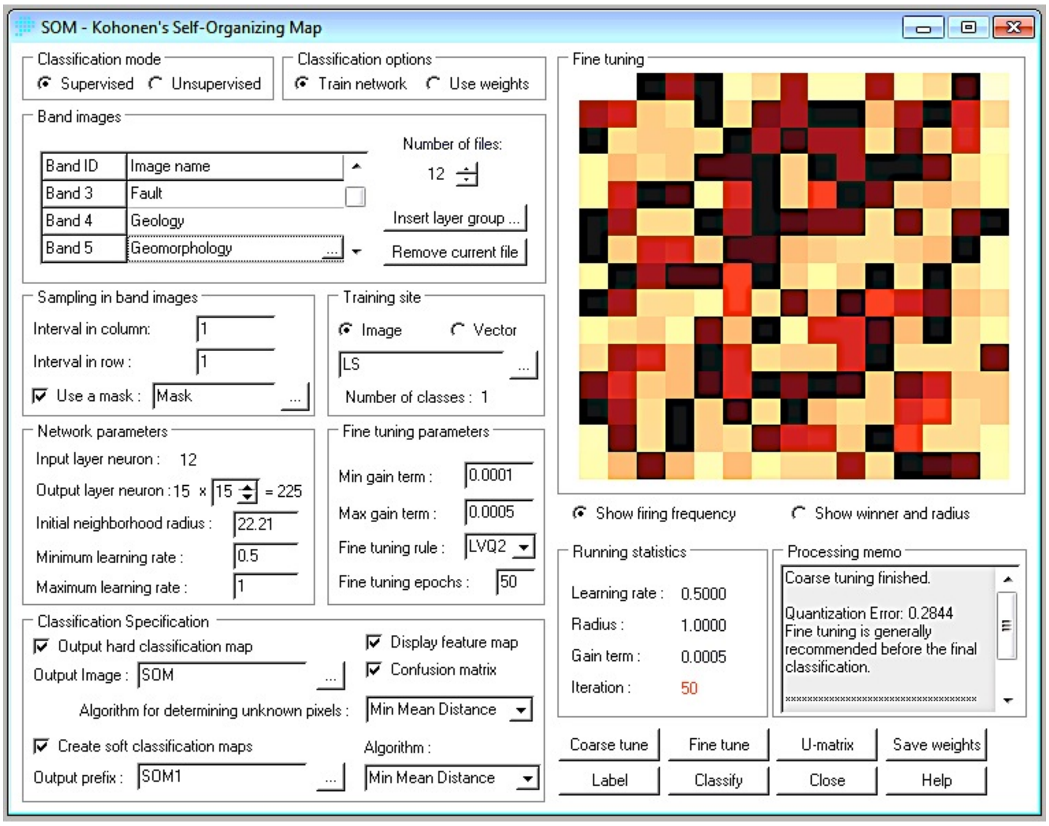

Figure 6. Screenshot for SOM calculation in IDRISI Selva software platform. Source: Bayes Ahmed, 2015.

For the given reference image, the following seven statistics are the same for all thresholds. The unit of each statistic is the proportion correct attributable to a combination of information of location and quantity.

No info of location and no info of quantity: N(n) = 0.5000; Perfect info of location and perfect info of quantity: P(p) = 1.0000; Perfect info of location and no info of quantity: P(n) = 0.5032.

No info of location and perfect info of quantity: N(p) = 0.9936.

No info of location and no info of quantity: PerfectChance = 0.5000.

No info of location and perfect info of quantity: PerfectQuantity = 0.4936.

Perfect info of location given no info of quantity: PerfectLocation = 0.0064.

The statistical details of the ROC findings are shown in Tables 3-5.

Table 3. Statistical measures for ROC.

No.	M(m)	N(m)	P(m)	M(p)	M(n)
1	0.9968	0.9968	0.9968	0.9936	0.5000
2	0.7529	0.7484	0.7532	0.9997	0.5030
3	0.5032	0.5000	0.5032	1.0000	0.5032
4	0.2532	0.2516	0.2532	1.0000	0.5032
5	0.0032	0.0032	0.0032	0.9936	0.5000

Source: Bayes Ahmed, 2015.

Table 4. Statistical measures for ROC.

No.	Kno	Klocation	Kquantity	Kstandard
1	0.994	0.000	1.006	0.000
2	0.506	0.946	0.503	0.018
3	0.006	1.000	-0.000	0.006
4	-0.494	1.000	-0.503	0.002
5	-0.994	0.000	-1.006	0.000

Source: Bayes Ahmed, 2015.

Table 5. Statistical measures for ROC.

No.	CorrectChance	CorrectQuantity	CorrectLocation	ErrorLocation	ErrorQuantity
1	0.500	0.497	0.000	0.000	0.003
2	0.500	0.248	0.005	0.000	0.247
3	0.500	0.000	0.003	0.000	0.497
4	0.252	0.000	0.002	0.000	0.747
5	0.003	0.000	0.000	0.000	0.997

Source: Bayes Ahmed, 2015.

The AUC is calculated 0.866 for SOM. On the other hand, using the similar datasets AUC for AHP, WLC, LR, and MLR methods were found as 88, 85.90, 74.90, and 90.40%, respectively (Ahmed, 2015a). Therefore, it can be stated that the results found in this article is statistically valid and representing results obtained from other similar analysis.

The SOM analysed some elements at risk layers like the building structure, road and drainage network maps. Other landslide hazard related layers were analysed like the slope, geomorphology, geology, soil moisture and lineaments etc. The community vulnerability is assessed based on the locational suitability of the settlements. For example, it is assumed that if a house is located in dangerous slope then the people residing in the house are more vulnerable to landslide hazards than a house located in a flat land. Considering these criteria, the SOM LSM map was classified into four different landslide hazard zones.

The accuracy of the SOM method can be increased if there is availability of high-resolution satellite and DEM images. This research is also lacking some important layers related to landslide causative factors such as structural map, and other detail soil/geological maps. In general, there is no such hypothesis like which method is better or worse than the other. The results can be different in another context or using better quality factor maps. Therefore, it would not be wise to conclude one particular method is applicable for a certain area in terms of landslide hazard/risk zoning. This is a trial and error process that depends on the quality and availability of datasets.

The final SOM image (Figure 7) was reclassified using 'Natural Breaks (Jenks)' method with 4 classes. Natural breaks classes are based on natural groupings inherent in the data. It identifies break points by picking the class breaks that best group similar values and maximize the differences between classes (Ahmed, 2015b). The SOM was represented with four categories of risk; namely- low risk (Green), medium risk (Yellow), high risk (Amber), and Very high risk (Red) as shown in Figure 7 (UDD, 2013, pp. 43).

5. Conclusions

Landslides are recognized as the third type of natural disaster in terms of the worldwide importance (van Westen et al., 2011, pp. 3-54). According to the world disasters report 2014, a total of 173 landslide disasters were reported worldwide (2004–2013) causing about 8739 human casualties and affecting 3.2 million people (IFRC, 2014). Although landslides are not considered as a severe disaster in the context of Bangladesh because calamities like floods, cyclones and sea-level rise are some prominent disasters where the Govt. is more concerned of taking mitigation measures.

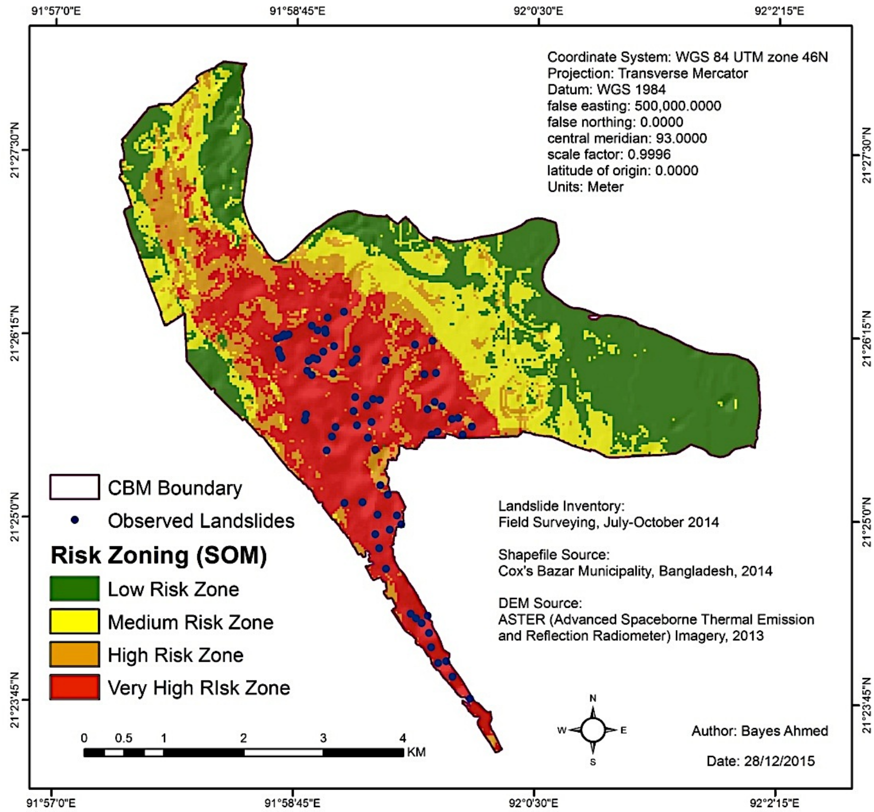

Figure 7. Landslide risk-zoning map applying SOM technique. Source: Bayes Ahmed, 2015.

The rate of urbanization is much higher in CHD due to excessive population growth, rural-urban migration and scarcity of flat land for developing residential settlements (Ahmed, 2015a). Therefore, landslides are going to be one the potential threats in the coming years for the hill district urban centres in Bangladesh. It should be also considered that earthquakes and climate change impacts on rainfall pattern could also trigger the landslide disasters in CHD. At this background, it is high time to produce the landslide hazard and risk zoning maps for CBM, which is the second biggest urban core in CHD after Chittagong Metropolitan Area. This paper has adopted the Kohonen's self-organizing map (SOM) neural network technique. The reasons behind choosing the SOM technique are statistical acceptance and higher accuracy. In addition, the SOM technique is first ever used for CBM in creating landslide risk zoning map.

To produce the SOM for landslide risk mapping a total of 12 factor maps (i.e. slope, land cover, geology, geomorphology, NDVI, soil moisture, rainfall pattern, and distance from-existing buildings; stream, road and drainage network, and faults and lineaments) are selected. A detail landslide inventory map is prepared for model validation and network training purpose. Then after fixing the parameter values the Kohonen's SOM technique was run in IDRISI Taiga software. The performance of the SOM method is validated using the ROC curve and the AUC value is calculated 86.60%. Then the SOM map is reclassified using natural breaks (Jenks) method in four different risk zones - very high, high, medium and low. The landslide risk-zoning map as produced by SOM technique is found scientifically significant and representing the ideal situation. The concerned urban planners, engineers and stakeholders can use this kind of risk-zoning maps for formulating policies related to landslide disaster risk reduction in the hill districts of Chittagong, Bangladesh.

6. Acknowledgements

Bayes Ahmed is a Commonwealth Scholar funded by the UK govt. The fieldwork in Cox's Bazar Municipality was partially funded by the Institute for Risk and Disaster Reduction, University College London (UCL), UK. The author is grateful to Professor Dr. David Alexander, Dr. Ilan Kelman, the research assistants from CUET URP, CBM officials and local people for helping in numerous occasions. Finally the author wants to thank the organizers of the 1st Bangladesh Planning Research Conference (BPRC) and the Department of Urban & Regional Planning, Jahangirnagar University, Bangladesh for creating the opportunity to publish this research article.

7. References

Ahmed, B. (2015a) Landslide susceptibility modelling applying user-defined weighting and data-driven statistical techniques in Cox's Bazar Municipality, Bangladesh. *Natural Hazards*, 79, pp. 1707–1737.

Ahmed, B. (2015b) Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides*, 12, pp. 1077-1095.

bdnews24. Available at: http://bdnews24.com/bangladesh/2015/06/26/10-killed-in-flash-flood-storm-landslides-in-coxs-bazar-bandarban [Accessed 28 December 2015].

Eastman, R. (2012) *IDRISI Selva help system*. Worcester: Clark Labs, Clark University.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P. (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 3, pp. 181–216.

IFRC. (2014) *World disasters report 2014—focus on culture and risk*. International Federation of Red Cross and Red Crescent Societies (IFRC), Geneva, Switzerland.

Kavzoglu, T., Sahin, E. K. and Colkesen, I. (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11, pp. 425–439.

Kohonen, T. (1990), The Self-Organizing Map. *Proceedings of the IEEE*, 78. pp. 1464-80.

Park, S., Choi, C., Kim, B. and Kim, J. (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environmental Earth Sciences*, 68, pp. 1443–1464.

Tso, B. and Mather, P. M. (2001) *Classification Methods for Remotely Sensed Data*. New York: Taylor and Francis.

UDD. (2013) *Guidelines for Mainstreaming Disaster Risk Reduction into Land Use Planning for Upazilas and Municipalities in Bangladesh*. Urban Development Directorate (UDD), Ministry of Housing and Public Works, Government of the People's Republic of Bangladesh and Asian Disaster Preparedness Center (ADPC), Thailand.

van Westen, C. J., Alkema, D., Damen, M. C. J., Kerle, N. and Kingma. N. C. (2011*) Multi-hazard risk assessment: Distance education course, Guide book*. United Nations University – ITC School on Disaster Geo- information Management (UNU-ITC DGIM).