

# **Development of GIS based Landslide Information System for the Region of East Sikkim**

**Shilpi Chakraborty**

Department of Computer Science and Engineering,  
Sikkim Manipal Institute of Technology,  
Sikkim, INDIA

**Ratika Pradhan**

Department of Computer Science and Engineering,  
Sikkim Manipal Institute of Technology,  
Sikkim, INDIA

## **ABSTRACT**

The use of remote sensing data, such as satellite imageries and aerial photos, allows us to map the variability of terrain properties, such as vegetation, water, geology, both in space and time. This paper suggests an integration of the classification of the landslides for the region of East Sikkim and the images are displayed through a web portal. The classification is done using k-means and k nearest neighbor classifier and the better one is selected to classify the landslides and then uploaded on a web portal.

## **Keywords**

Image classification, k-means, training set, web portal, k-nearest neighbor.

## **1. INTRODUCTION**

Landslides are the downslide movement of soil, debris or rocks, resulting from natural causes, vibrations, overburden of rock material, removal of lateral supports, change in the water content of rock or soil bodies, blocked drainage etc. The fragile nature of the rocks forming the mountains, along with climatic condition and various anthropogenic activities has made the state of Sikkim vulnerable to the vagaries of nature. Landslide is the most costly and damaging natural hazards in the mountainous region. These landslides are triggered mainly by rainfall or earthquakes. The growing population and expansion of settlements over these hazardous areas increased the impact of natural disasters in most of these areas. The landslide hazards, in general cannot be completely prevented. However, the intensity and severity of impacts of the hazards can be minimized if the problem is recognized before the development activity. Hence, there arises a need to identify the unstable slopes using the landslide hazard zonation. The landslide hazard zonation (LHZ) of an area aims at identifying the landslide potential zones and ranking them in order of the degree of hazard from landslides. LHZ mapping involves a detailed assessment and analysis of the past occurrences of landslides in conditions of their location, size and incidence with respect to various geo-environmental factors that cause landslides and mass movements. LHZ map separates and draw out varying degrees of predictable slope stability. The map has an inbuilt factor of forecasting and is of probabilistic nature. Depending upon the methodology adopted and the comprehensiveness of the input data used, a LHZ map is able to provide help concerning some or all the following individual factor maps like landslide location, slope steepness, landuse/ landcover, geology or lithology, density of drainages, rainfall etc. A huge quantity of data on large slope areas are collected, stored, sorted and evaluated. Then the level of risk sliding has to be assessed and zonation maps prepared. The use of aerial photographs, satellite images and adoption of remote sensing techniques helps in the collection of data. For

storage, retrieval and analysis, adoption of computerized techniques would be useful.[1]

Remote sensing is the art and science of making measurements of the earth using sensors on airplanes or satellites. These sensors collect data in the form of images and provide specialized capabilities for manipulating, analyzing, and visualizing those images. Remote sensed imagery is integrated within a GIS. A geographic information system is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The acronym GIS stands for Geographical Information Science or Geospatial Information Studies to refer to the academic discipline or career of working with geographic information systems. A GIS is a computer-based tool for mapping and analyzing feature events on earth. GIS technology integrates common database operations, such as query and statistical analysis, with maps. GIS manages location-based information and provides tools for display and analysis of various statistics, including population characteristics, economic development opportunities, and vegetation types. GIS allows linking of databases and maps to create dynamic displays. It also provides tools to visualize, query, and overlay those databases in ways not possible with traditional spreadsheets. These abilities distinguish GIS from other information systems, and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. The enormous demand for the storage, analysis and display of complex and voluminous data has led, in recent years, to the use of Geographic Information Systems for effective data handling and also for analyzing and geographically transferring the information around the world. Remote Sensing and GIS is a very useful tool to complement conventional methods involved in Disaster Management. Mitigation of natural disaster management can be successful only when detailed knowledge is obtained about the expected frequency, character, and magnitude of hazard events in an area. Although, natural disasters have shown in the last few decades a drastic increase in magnitude and frequency, it can as be observed that there has been a dramatic improvement in technical capabilities to mitigate them. The use of remote sensing data, such as satellite imageries and aerial photos, allows us to map the variability of terrain properties, such as vegetation, water, geology, both in space and time. Satellite images give a synoptic overview and provide very useful environmental information, for a wide range of scales, from entire continents to detail of a few meters. Many types of disaster, such as floods, droughts, earthquakes, etc. will have certain precursors that satellite can detect. Remote sensing also allows monitoring the event as it occurs.

## 2. RELATED WORK

In a paper “Color Image Clustering using k-means” by K.Sneha Silva et al, the clustering is performed in images based on their color. This is done by quantizing the color distribution into color histograms. RGB color components are taken from each and every image. Then, the mean values of Red, Green, and Blue components of target images are calculated and stored in the database. Based on the RGB component mean values, the images are clustered as Red, Green and Blue major component categories. These three mean values for each image are stored and considered as features [2]. A method of comparison of clusters using an improvised k-means and neural networks were performed in the paper “Land-cover Classification and Mapping for Eastern Himalayan State Sikkim by Ratika Pradhan et al.[3]A paper written by D S Guru et al “Texture Features and KNN in Classification of Flower Images”, an unknown flower is queried to K-nearest neighbor classifier after performing the training phase and a classification phase. In training phase, from a given set of training images the texture features (GLCM / Gabor / Combination) are extracted and used to train the system using the K-nearest neighbor classifier. In classification phase a given test flower image is segmented and then the above mentioned texture features are extracted for classification. These features are queried to K-nearest neighbor classifier to label an unknown flower.[4]

In a paper “Color Based Image Segmentation Using K-Means Clustering” by Anil Z. Chitade et al presents a novel image segmentation based on color features with K-means clustering unsupervised algorithm. In this no training data was used. The entire work is divided into two stages. First enhancement of color separation of satellite image using decorrelation stretching is carried out and then the regions are grouped into a set of five classes using K-means clustering algorithm. Using this two step process, it is possible to reduce the computational cost avoiding feature calculation for every pixel in the image. Although the color is not frequently used for image segmentation, it gives a high discriminative power of regions present in the image. [5]

## 3. STUDY AREA

Sikkim is 22nd state of India and is mountainous state that shares its boundary with three sovereign nations, Nepal in west, Bhutan in east and Tibet, China in north. The state shares its southern boundary with Darjeeling district of West Bengal, India. The state is situated between 27004’46” and 28007’48” north latitudes and 88000’58” and 88055’25” east longitude. The state has total geographical area of 7096 sq km. The study area is considered as the region of East Sikkim, 88.30 east and 27.30 north 88.58925 east and 27.3223 north.



Fig 1: Map of East Sikkim

## 4. K-MEANS CLASSIFICATION

K-means algorithm takes the input parameter, k, and partitions a set of ‘n’ objects into k-clusters so that the resulting intra-cluster similarity is high but the inter-cluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster. It randomly selects ‘k’ of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the objects and the cluster mean. It then computes the new mean for each cluster. This process iterates until the criterion function converges. The square-error criterion is defined as,

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

Where,

‘E’ is the sum of the square error for all objects in the data set, ‘p’ is the point in space representing a given object,  $m_i$  is the mean of cluster  $C_i$  ( both p and  $m_i$  are multi-dimensional)

The steps involved are as follows:

- (i) Place ‘k’ points into the space represented by the objects that are being clustered. These points represent initial group centroids.
- (ii) Assign each object to the group that has the closest centroid.
- (iii) When all the objects have been assigned, recalculate the positions of the k-centroids.
- (iv) Repeat (ii) and (iii) until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

## 5. K NEAREST NEIGHBOR CLASSIFICATION

The KNN algorithm was originally suggested by Cover and Hart. K-nearest neighbor algorithm (KNN) is part of supervised learning that has been used in many applications in the field of data mining, statistical pattern recognition etc. [17] In pattern recognition, the k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. KNN is a method for classifying objects based on closest training examples in the feature space. An object is classified by a majority vote of its neighbors. K is always a positive integer. The neighbors are taken from a set of objects for which the correct classification is known. It is usual to use the Euclidean distance, though other distance measures such as the Manhattan distance can also be used.

The steps of the algorithm are as follows:

- (i) Determine the parameter K = number of nearest neighbors beforehand. This value is user-defined.
- (ii) Calculate the distance between the query-instance and all the training samples. Here Euclidean distance is used.

(iii) Sort the distances for all the training samples and determine the nearest neighbor based on the K-th minimum distance.

(iv) Since this is supervised learning, get all the categories of the training data for the sorted value which fall under K.

(v) Use the majority of nearest neighbors as the prediction value.

## 6. RESULTS AND DISCUSSION

The classified images are as follows:

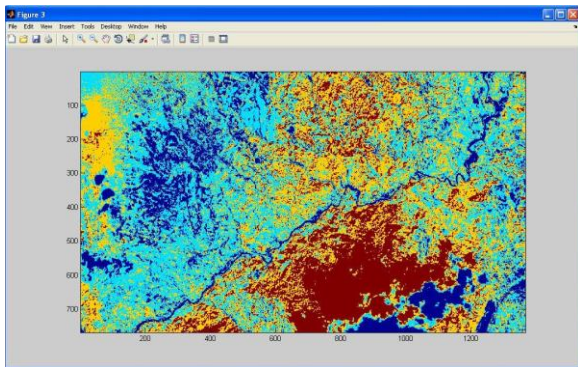


Fig 2: Classified satellite image using k-means

Legend:

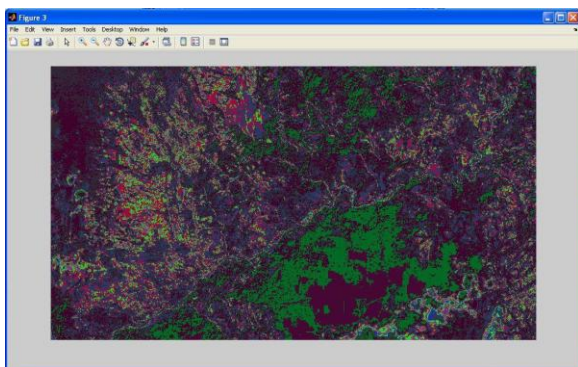
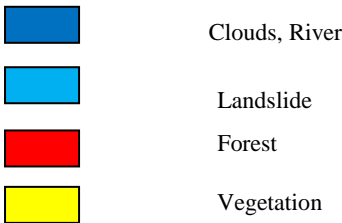
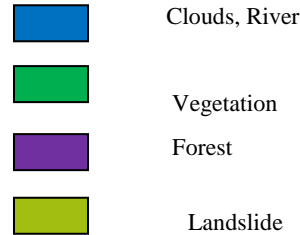


Fig 3: Classified satellite image using k nearest neighbor classifier

Legend:

The number of clusters for each algorithm is considered to be  $K=5$ . The graph representing the number of pixels can be shown as below:



The snapshots of the created web portal are as follows:



Fig 4: Album with uploaded image

The snapshot of the database tables are as follows:

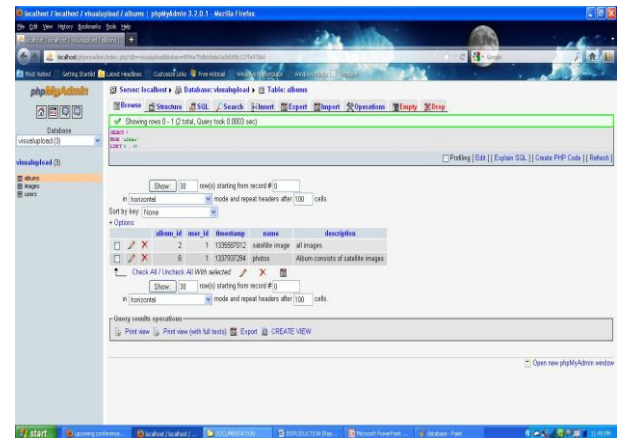


Fig 5: Table created to store the album details

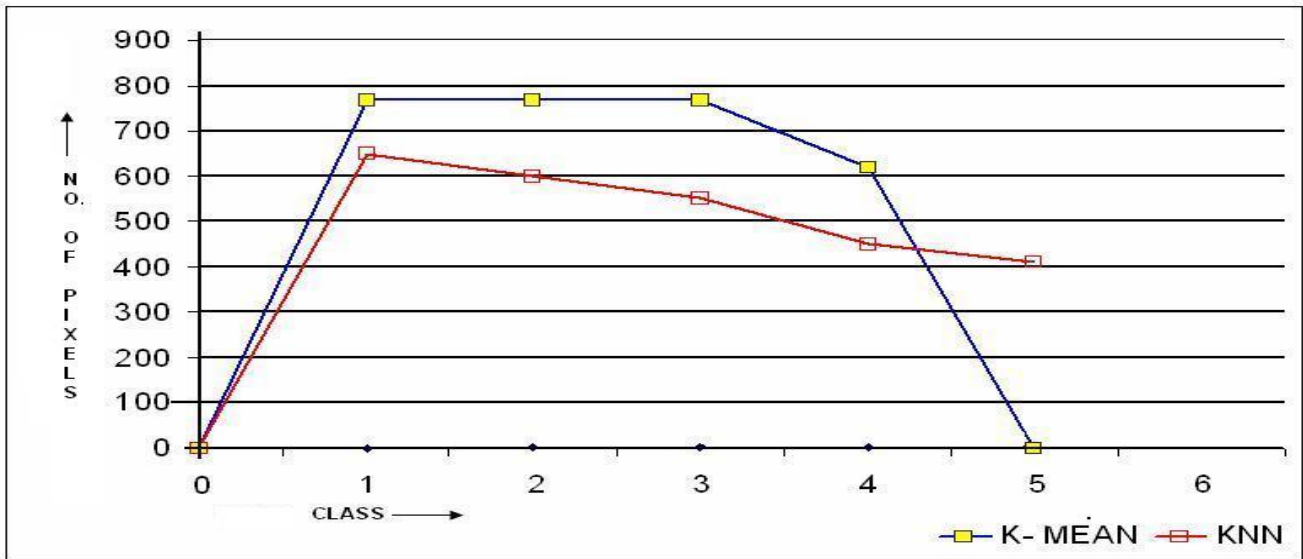


Fig 6: Graph plotted to show the number of pixels detected by k-means and knn algorithms.

A confusion matrix for the supervised classification is drawn based on the identification of Landslide, clouds, forest and vegetation from the training set.

Table 1. Confusion Matrix

Ground Class	Landslide	Clouds	Forest	Vegetation	Omissions	Commissions	Map Accuracy
Landslide	3377	73	53	65	6.51%	5.36%	88.13%
Clouds	50	7352	44	34	2.65%	1.72%	95.63%
Forest	100	52	6976	0	2.14%	2.14%	95.72%
Vegetation	85	75	55	5801	1.68%	3.58%	94.74%

For unsupervised classification method of classifying image observer is the best judge; however one can visit the place physically of ground truth collection. For supervised classification method the confusion matrix is shown above in the table which gives overall classification accuracy of 97.164%. From the confusion matrix shown in the table it is clear that supervised KNN gives best result for almost all the classes except for the landslide, as some the pixel values are mis-classified to the other forest classes i.e. vegetation and forest.

## 7. CONCLUSION

A web portal is built using PHP, HTML and JavaScript. The styling is done using CSS. The web portal has the features of creating a new user and providing a login page to the existing user with a valid login id and password. It can create multiple albums and upload images to it. It also provides the feature of viewing the album and deleting images and albums from a particular user account. The users are the authority who has the capacity to upload or delete images. The clients can only view the images when the web portal is uploaded into the internet.

The classification methods k-means and k-nearest neighbor classifies the regions from a satellite image. The k-means algorithm is fast compared to KNN but has a problem of overlapping clusters and empty clusters. Whereas the KNN is more accurate than k-means but it takes long time duration to complete its execution. The time taken to achieve the output increases with the increase in size of the training set.

The report is generated with a specific format depicting the location, date and time of occurrence. Landslides are categorically evaluated and the report is forwarded to the concerned authority.

The algorithm works properly for the class clouds, forest and vegetation but shows 88.13% accuracy for the class landslide. The performance can improve with increasing the size of the training data set which will help the algorithm to classify the landslides more accurately.

The future scope of the project is that the web portal can be enhanced to a client-server system where the clients can send e-mails to the server and the server can respond to the query of the clients after performing the required search in its database and generate an e-mail to the client stating the result if found. Also the project can be enhanced to the entire state of Sikkim wherein a huge database will be required.

## 8. REFERENCES

- [1] S.Evany Nithya, P. Rajesh Prasanna, “An Integrated Approach With GIS And Remote Sensing Technique For Landslide Hazard Zonation”, International Journal Of Geomatics And Geosciences Volume 1, No 1, 2010, ISSN 0976 – 4380
- [2] K. Sneha Silvia, Y. Vamsidhar, G. Sudhakar, “Colour Image Clustering using K-Means”, IJCST Vol. 2, SP 1, December 2011, ISSN 0976-8491(Online) | ISSN : 2229-4333(Print).

- [3] Ratika Pradhan, Mohan P. Pradhan, Ashish Bhusan, Ronak. K. Pradhan, M. K. Ghose, “Land-cover Classification and Mapping for Eastern Himalayan State Sikkim”, *Journal Of Computing*, Volume 2, Issue 3, March 2010, ISSN 2151-9617
- [4] D S Guru, Y. H. Sharath, S. Manjunath, “Texture Features and KNN in Classification of Flower Images”, *IJCA* Special Issue on “Recent Trends in Image Processing and Pattern Recognition”, RTIPPR, 2010.
- [5] Anil Z Chitade, Dr. S.K. Katiyar, Colour Based Image Segmentation Using K- Means Clustering, *International Journal Of Engineering Science And Technology* Vol. 2(10), 2010, 5319-5325.