Indian Journal of Science and Technology, *Vol 9(21), DOI: 10.17485/ijst/2016/v9i21/95275, June 2016*
*ISSN (Print) : 0974-6846*
*ISSN (Online) : 0974-5645*

# Landslide Prediction with Rainfall Analysis using Support Vector Machine

**Neenu Rachel[1*] and M. Lakshmi[2]**

[1]Department of Computer Science and Engineering, Sathyabama University, Chennai - 600119, Tamil Nadu, India; neenurachel_28@yahoo.com
[2]Faculty of Computing, Sathyabama University, Chennai - 600119, Tamil Nadu, India; laks@icadsindia.com

## Abstract

**Objective:** The paper aims in presenting a prediction model by using Support Vector Machine (SVM) technique which is meant to possess a strong capability to predict landslides by forecasting rainfall dataset using BigData concept. **Methods:** The dataset has been taken for the Cherapunjee region which receives the highest intensity of rainfall in India. The aim is to predict the landslide occurrence and classify the risk level associated with the landslide. To improve the reliability in landslide prediction, the proposed model uses pre-processing for removing null values in the dataset. After getting the pre-processed dataset, it will apply normalization, then SVM training and finally the Testing process. Thus the Support Vector Machine concept proved to exhibit a large degree of flexibility in handling tasks of varied complexities because of the non-linear boundary functions. **Findings:** The study concludes that SVM proved to be an efficient technique to forecast the landslides by predicting the rainfall in advance. The comparative results of SVM in regard with Artificial Neural Networks were proven. The study has been done specifically for Cherrapunjee region and can be implemented for any landslide prone area. **Novelty/Improvement:** Researchers worldwide are having a great pace to develop early prediction mechanisms for natural hazards. The study uses Radial Basis Function as an initial parameter for predicting the risk level classification of landslide. The novelty is in providing an initial selection of the kernel parameter in order to save the time on finding the best parameters.

**Keywords:** BigData, Hadoop, Rainfall Data, SVM

## 1. Introduction

The rainfall induced hazard has seriously influenced the sustainable development in many parts of the country. The effect of damages and losses mainly incurred regularly as there has been very less considerations of the potential problems in the planning of land usage and slope stability management. The landslide term can be describes as the movement of the slope downwards along with debris due to a sudden outbreak in the rainfall leading to mass flow of land and other substances. Landslide occurrence can be defined as a function of both temporal and spatial variability of the site scenarios[1].

The regional precipitation induced landslides in Uttharakhand, Meghalaya, Sikkim and the Western Ghats[2-4]. A statistical approach to slope stability analysis with the strength reduction methodology is carried for the Nilgiri region[4]. The authors have proposed for a modular ANN mechanism which is coupled with the data-pre-processing strategies. It aims in improving the accuracy of rainfall forecasting. MANN is comprised of 3 local models which are part of the 3 subsets being clustered by the Fuzzy C-Means (FCM) clustering technique.

The recent occurrence with the highest devastation which occurred in Utharakhand in 2013 caused by heavy rainfall took away the lives of more than thousand lives. Landslide phenomenon presents a significant pronouncement during the monsoon period[5].

The researchers have employed various models which include both empirical as well as mathematical concepts such as ANN to predict rainfall intensity and access the risk of landslide occurrence in comparison with the rainfall threshold. But here the prediction accuracy is

**Author for correspondence*

observed to be low and computational time is high.

With the advancement of computational efficiency which when combined with several sophisticated statistical approaches[6], various techniques have been used for the prediction of rainfall induced landslides. The approach of a linear system is quite complex when climatic phenomenon are considered. The ANN which was invented in 1964 was made use in landslide prediction using the daily rainfall analysis. The back propagation algorithm was made used for the mapping of the input to the output with the appropriate back propagation of errors.

Several risk analysis are performed using the remote sensing data. The authors have used the Digital Elevation Model technique [7] to explore the most hazardous zones of the landslides area. The DEM has been used to calculate the various factors such as slope aspects, slope angle and the slope curvature.

An extensive study on the geological, geotechnical and the geophysical investigations[8] were carried out to study the major cause of the landslide activation in the Lanta Khola region of the Sikkim Himalayas. The authors have strengthened their model by proving that the sliding and the debris generation are initiated in the zone only after the heavy rainfall[9].

An Artificial Neural Network model which is categorized on frequency analysis using the Fast Fourier Transform (FFT) technique[10] has been used. The study also has examined the roles of Feed Forward Neural Network (FFNN) for the rainfall prediction. A detailed accuracy classification of the prediction of landslide occurrence has been depicted using SVM and ANN technique

Thus several mathematical and empirical models have been extracted to achieve different precision in several regions across the world. Many studies are also concentrated on both the temporal as well as spatial parameters. The paper exhorts a cluster validity index for rough fuzzy c-means[11] clustering algorithm called rough fuzzy Bayesian like validation method which roots on probabilistic metric.

The very motive is to develop a landslide prediction model and generating the risk level associated with the landslide. This can increase the preparedness level to reduce the severity of the disaster thereby having a control on the human loss and also the infrastructural damages. As far as this area of study is concerned, there is no landslide prediction system in this region. This paper employs the Support Vector Machine Concept to develop an early warning approach to the landslide. The risk level (high, medium, low) is also evaluated in this paper.

## 2. Methodology

### 2.1 Research Motivation

The existing system uses artificial neural network to predict the rainfall intensity, one day in advance. The daily threshold values are used for comparison with the rainfall intensity to forecast the occurrence of landslide.ANN removes the complex parameters in the physical process and thus considered to be black-box. The enhancement of performance and volume management are the 2 major goals of research works in the domain of Bigdata.

### 2.2 Support Vector Machine

Support Vector Machine(SVM) originally developed by Vapnik, in 2005 is considered to be a new generation of learning algorithms. The Support Vector Machine concept has been used in accordance with the constructive machine learning procedure based on the statistical learning theory. It provides non-linear solutions to regression as well as classification problems by transforming the inputs which occupy a large dimension space. In this paper, SVM model has been used for both training and testing process.

### 2.3 Application of SVM in the Study Area

The forecasting of landslide has been dealt using factors which affects the landslide occurrence. We have used the training of data samples to determine several classes for prediction and to remove the null values. Also, the testing process has been done to improve efficiency and accuracy.

#### *2.3.1 Dataset*

The Cherrapunjee region of Meghalaya has been selected for the study which has the maximum intensity of rainfall in the country. The dataset comprises of 7 years daily recorded data for rainfall from 2009 to 201512 as in figure 1.

#### *2.3.2 Selection of Parameters*

The several other parameters in concern with the rainfall data included the maximum and minimum temperature, air pressures during morning and evening and wind speed. A proper class based approach has been taken to distinguish the several parameters.

To build a SVM model with much efficiency and accuracy, the parameters (C and γ) of the prediction model need to be chosen in advance. The parameter C is used to determine the trade-off cost in between for minimizing the error in training and also the complexity of the SVM model. The predictive accuracy of the training sample will be higher if the value of C increases. This however might cause an over-training issue. The parameter used in the Radial Basis Function determines a non-linear mapping from the input space to the high-dimensional feature space. Shape of RBF function is determined by the value of γ. Therefore, the parameters (C and γ) have a strong influence on the generalization performance and efficiency of SVM model.

### *2.3.3 Computing Procedure*

The procedure of computing the landslide prediction model is accomplished using the following steps:

- Transform data to remove null values
- Perform scaling on the transformed data
- Consider the RBF kernel $K(X_i,X_j)=\exp(-\gamma\ ||x_i-x_j||^2)$
- Find the best parameter C and γ using cross-validation
- Train the whole training set by using the best parameter C and γ
- Test the data

## 3. Algorithm

### 3.1 Landslide Prediction Algorithm

The landslide prediction algorithm has three phases, the algorithm for predicting landslide is briefly explained below

***Algorithm 1: Landslide_prediction_Algorithm***

*Input:* Climatic data from the IndianMeteorological Observatory

*Output:* Display the landslide occurrence dates and classify the risk level of it.

The main Detection procedure for landslide prediction algorithm is given as:

1. Start
2. {
3. Rainfall dataset as an input
4. Initial input size
5. Set threshold value of Rainfall
6. initialize threshold to zero
7. {
8. Initialize the input format
9. Training of the rainfall dataset with SVM
10. Calculate the error rate of trained dataset
11. Testing of the dataset with SVM
12. Normalization of Trained dataset
13. Set maximum and minimum fraction
14. Start the Yearly evaluation of the landslide
15. End the yearly evaluation of the landslide
16. {
17. Compute Threshold
18. Predict the risk level of land slide
19. }
20. }
21. }
22. End

The proposed work has been implemented as in 4 modules. They are described as follows:

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | |
| 2 | Date | Min Temp (Celsius) | Max Temp (Celsius) | Morn. Air Pressure (mb) | Even. Air Pressure (mb) | Rainfall (mm) | Annual Rainfall (mm) | Wind Speed (k/h) |
| 3 | 17/08/2015 | 18.3 | NA | 866.9 | NA | 162.2 | 8388.2 | NA |
| 4 | 16/08/2015 | 18.3 | 21 | 867.8 | 865.4 | 194.9 | 8226 | NA |
| 5 | 15/08/2015 | 18 | 21.2 | 867.8 | 865.3 | 191.2 | 8031.1 | NA |
| 6 | 14/08/2015 | 17.8 | 22.9 | 866.2 | 865 | 101.8 | 7839.9 | NA |
| 7 | 13/08/2015 | 18.5 | 24.2 | 867.8 | 864.7 | 4 | 7738.1 | NA |
| 8 | 08-12-2015 | 19 | 25 | 868.4 | 865.6 | 0 | 7734.1 | NA |
| 9 | 08-11-2015 | 19.5 | 26.5 | 868.1 | 865.1 | 20.6 | 7734.1 | NA |
| 10 | 08-10-2015 | 19.4 | 21.8 | 865.4 | 862.9 | 1.8 | 7713.5 | NA |
| 11 | 08-09-2015 | 19.1 | 26.6 | 864.2 | 861.4 | 0 | 7711.7 | NA |
| 12 | 08-08-2015 | 18.4 | 25.9 | 866 | 863.1 | 0 | 7711.7 | NA |
| 13 | 08-07-2015 | 17.5 | 26 | 864.8 | 863.2 | 0.8 | 7711.7 | NA |
| 14 | 08-06-2015 | 17.1 | 24.3 | 864 | 860.2 | 6.8 | 7710.9 | NA |
| 15 | 08-05-2015 | 17.5 | 20.8 | 864.5 | 861.7 | 58.6 | 7704.1 | NA |
| 16 | 08-04-2015 | 18.3 | 22 | 864.8 | 863.2 | 4 | 7645.5 | NA |
| 17 | 08-03-2015 | 17.9 | 24.2 | 866.9 | 864 | 0 | 7641.5 | NA |
| 18 | 08-02-2015 | 17.4 | 23.2 | 867.7 | 864.8 | 27.9 | 7641.5 | NA |
| 19 | 08-01-2015 | 17.4 | 19.3 | 864.7 | 864.8 | 85.4 | 7613.6 | NA |
| 20 | 31/07/2015 | 17.3 | 20.2 | 859.5 | 860.1 | 10.4 | 7528.2 | NA |
| 21 | 30/07/2015 | 21.2 | 26.9 | 859.9 | 857.7 | 0 | 7517.8 | NA |
| 22 | 29/07/2015 | 20.6 | 29.2 | 861.6 | 859.1 | 0 | 7517.8 | NA |
| 23 | 28/07/2015 | 19.4 | 27.2 | 862.9 | 860.4 | 0 | 7517.8 | NA |
| 24 | 27/07/2015 | 19 | 26.8 | 862.7 | 861.1 | 0.1 | 7517.8 | NA |
| 25 | 26/07/2015 | 18.9 | 26.6 | 863.3 | 861.9 | 0 | 7517.7 | NA |
| 26 | 25/07/2015 | 16.6 | 24.6 | 864.5 | 862.7 | 74.7 | 7517.7 | NA |
| 27 | 24/07/2015 | 17.9 | 22.2 | 867.3 | 864.3 | 258.6 | 7443 | NA |
| 28 | 23/07/2015 | 17.9 | 21.8 | 867.7 | 865.5 | 51.2 | 7184.4 | NA |

**Figure 1.** Sample dataset.

- Landslide Dataset Preparation and Pre-processing
- Normalization and SVM Training
- Landslide Prediction using SVM
- Landslide Stage Identification

### *3.1.1 Landslide Dataset Preparation and Pre-processing*

The Prediction of landslide based on the previous year's data and records. In first step, for predicting the landslide, need to collect all the previous record for pre-processing function, which makes a specific result for the future Prediction. The pre-processing method removes the null value.

### *3.1.2 Normalization and SVM Training*

The Support Vector Machine (SVM) trains the dataset by using Radial Basis Function. Radial Basis Function is one of the major kernel functions that should be utilized for SVM training process. The trained data is classified and tested by SVM. Classification is the process where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training.

### *3.1.3 Landslide Prediction using SVM*

The users make a request for the testing process by Support Vector Machine technique. The testing is essential process for predicting the landslide based on the classification. To make the classification results comparable and for exhaustive data analysis, we have used leave one out classification method for the SVM classifier. In classification, SVM efficiently classifies the landslide dataset.

### *3.1.4 Landslide Stage Identification*

Predicted landslide image taken as an input for this module. In this module, analyse the risk level of future prediction landslide dataset. Consider three types of risk level. The climatic threshold based mechanism attempts to classify the dataset into 3 stages: low, medium, high. Finally display the risk level of landslide such as (low, medium and high).

## 4. Architecture

The parameters taken as input includes X1, X2,…$X_n$ where n represents the number of input nodes in fig 2.The Rainfall Threshold for the day is represented as $R_t(t)$ and the Rainfall Predicted for the day as $R_o(t)$.

## 5. Design and implementation:

### 5.1 Design

The design concept of the landslide prediction model has been explained in accordance with the algorithms mentioned above. Initially Collect Landslide Datasets from the user and store the datasets into database. Then apply

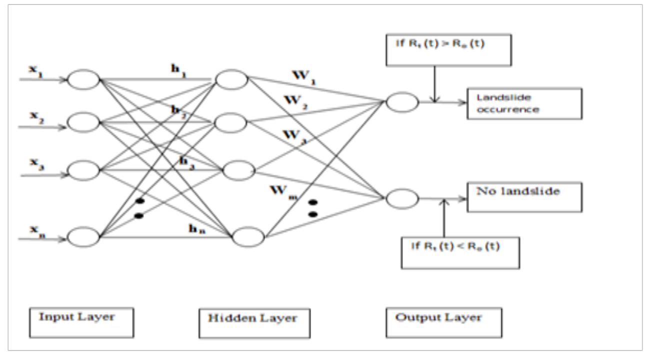


**Figure 2.** SVM Architecture

pre-processing to replace all the null values (only applicable for numerical dataset).Pre-processing reduce the time complexity. Then pre-processed data were taken as an input for training process.

Apply normalization process for normalizing the pre-processed dataset. Radial Basis Function is one of the major kernel functions is utilized for SVM training process.

In the testing phase, Classification process is used where a given test sample is assigned a class on the basis of knowledge gained by the classifier during training. Then proceed by taking the query dataset for testing phase. Apply pre-processing for input query dataset.

Support Vector Machine compares the query dataset with trained dataset. It displays the predicted status

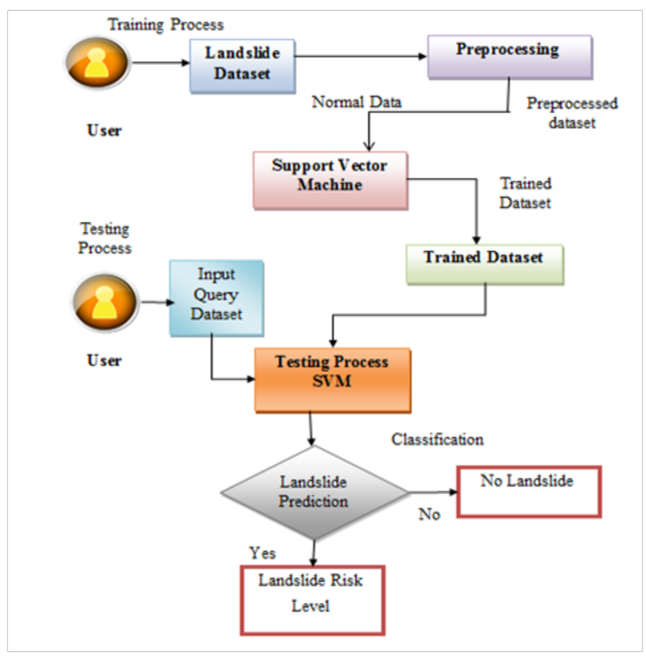


**Figure 3.** Flow diagram of Landslide prediction.

(landslide or not).Apply climatic threshold based prediction mechanism if the classification result is relevant to landslide. Finally display the risk level of landslide such as (low, medium and high) as in figure 3.

### 5.2 Implementation

The proposed algorithm is implemented using hadoop with java language. The algorithm is implemented considering the nature of the spatial database. The details of the nodes are stored in the hive database. Hadoop is an open-source software framework that supports data-intensive distributed applications to facilitate the storage and processing of large and rapidly growing data sets. Since it is highly scalable, they are used with increased no of attributes.

## 6. Results

The above graph figure 4 is showing the prediction of rainfall comparison with real, ANN and proposed technique SVM. The result is showing that the proposed technique SVM is better than ANN and it produces more relevant and approximate result.

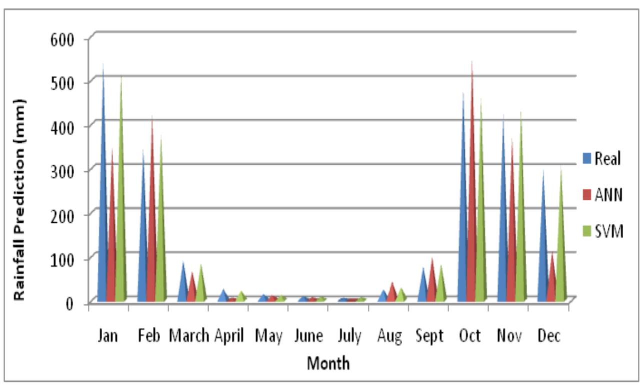


**Figure 4.** Comparison with Real, ANN, SVM.

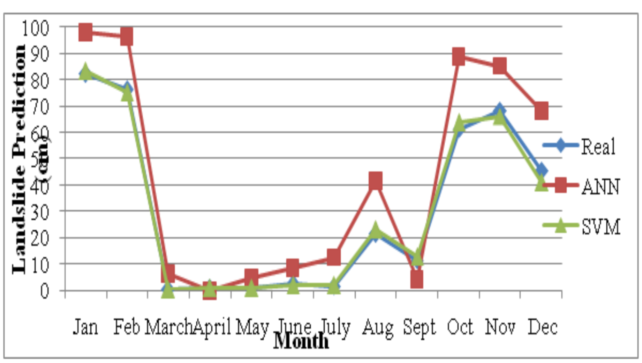


**Figure 5.** Comparison with Real, ANN, SVM.

The above graph fig 5 is showing the landslide prediction accuracy within the rainfall condition, the proposed technique SVM is producing more relevant and close to exact or real prediction of the landslide.From subsequent above tasks, the training and the testing samples for the study area was accomplished. Different parameters in training process were inputted. Construction of the prediction model was done with the best parameters and thus the model was used to pick up the support vectors in the samples. We verified from various literature surveys that using Library Support Vector Machine (LIBSVM) for training the dataset sample would provide better results.

## 7. Conclusion and Future Work

The natural hazards are beyond human control but their destruction can be reduced if prediction mechanisms are carried out in advance. Researchers worldwide are having a great pace to develop early prediction mechanisms for such natural hazards. It was hard to use traditional mathematical methods for analyzing. Thus SVM helped in getting non-linear relationships using the historical data.

Thereby, the study concludes that SVM proved to be an efficient technique to forecast the landslides by predicting the rainfall in advance. The study has been done specifically for Cherrapunjee region and can be implemented for any landslide prone area

The application of the landslide prediction model is an exploration always and it includes many things to be researched further. The future scope of the work is in providing an initial selection of the kernel parameter in order to save the time on finding the best parameters.

## 8. Acknowledgement

The access to rainfall dataset of Cherrapunjee datasets from the Indian Meteorological Observatory has been acknowledged. We are also thankful to the Sathyabama University for encouragement.

## 9. References

1. Renuga Devi S, Agarwal P, Venkatesh C, Arulmozhivarman P. Daily Rainfall Forecasting using Artificial Neural Networks for Early Warning of Landslides, IEEE, International Conference on Advances in Computing, Communications and Informatics. 2014 Sep. p. 2218–24.

2. Anbarasu K, Sengupta A, Gupta S, Sharma SP. Mechanism of activation of the Lanta Khola landslide in Sikkim Himalayas, Landslides. 2010 Jun; 7(2):135–47.
3. Sengupta A, Gupta S, Anbarasu K. Rainfall thresholds for the initiation of landslide at Lanta Khola in north Sikkim, India, National Hazards. 2010 Jan, 52(1):31–42.
4. Chandrasekaran SS, Elayaraja S, Renugadevi S. Damages to transport facilities by rainfall induced landslides during November Nilgiris, India Proceedings of the Second World Landslide Forum, Rome, 2011Oct.
5. National disaster management guidelines – Management of landslide and snow avalanches. A publication of the national disaster management authority, Government of India, 2009 Jun.
6. Zhao X, Ji M, Cui X. Research On Landslide Prediction Model Based On Support Vector Model, The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences. 2010 Jun; 3:540–4.
7. Pradhan B, Lee S. Landslide risk analysis using artificial neural network model focusing on different training sites. International Journal of Physical Sciences. 2009 Feb; 4 (1):1–15.
8. Anbarasu K, Sengupta A, Gupta S, Sharma SP. Mechanism of activation of the Lanta Khola landslide in Sikkim Himalayas, Springer-Verlag. 2010 Jun; 7 (2):135–47.
9. Mahajan S, Mazumdar H. Rainfall Prediction using Neural Net based Frequency Analysis Approach. International Journal of Computer Applications. 2013 Dec; 84(9):7–11.
10. Yao X, Dai FC. Support vector machine modelling of landslide susceptibility using a GIS: A case study, The Geological Society of London, IAEG. 2006; 1–12.
11. Revathy S, Parvaathavarthini B, Rajathi S. Futuristic Validation Method for Rough Fuzzy Clustering. Indian Journal of Science and Technology. 2015 Jan; 8(2):120–7.
12. Cherrapunjee dataset. Available from: http://www.cherrapunjee.com/weather-info/daily-weather-data-2009-2013/ Date accessed: 11/09/2015