

# Prediction of Shallow Landslide prone regions in Undulating Terrains

Ramachandra T. V.<sup>1,2,3,\*</sup>, Bharath H. Aithal,<sup>1,2</sup> Uttam Kumar<sup>1</sup> and Joshi N. V.<sup>1</sup>

1. Energy & Wetlands Research Group, Center for Ecological Sciences [CES], Indian Institute of Science, Bangalore, INDIA

2. Centre for Sustainable Technologies (ASTRA), Indian Institute of Science, Bangalore, Karnataka, 560 012, INDIA

3. Centre for infrastructure, Sustainable Transportation and Urban Planning [CISTUP], Indian Institute of Science, INDIA

\*cestvr@ces.iisc.ernet.in

## Abstract

*Genetic Algorithm for Rule-set Prediction (GARP) and Support Vector Machine (SVM) with free and open source software (FOSS) – Open Modeller were used to model the probable landslide occurrence points. Environmental layers such as aspect, digital elevation, flow accumulation, flow direction, slope, land cover, compound topographic index and precipitation have been used in modeling. Simulated output of these techniques is validated with the actual landslide occurrence points, which showed 92% (GARP) and 96% (SVM) accuracy considering precipitation in the wettest month and 91% and 94% accuracy considering precipitation in the wettest quarter of the year.*

**Keywords:** Landslide, Genetic algorithm, Support Vector Machine.

## Introduction

Landslides occur due to various triggering factors such as heavy rainfall, earthquakes, changes in land cover etc. These hydro-geomorphic events depend on the combination of i) predisposing factors (e.g. lithology and morphology), ii) triggering factors (e.g. excessive and intense precipitations) and iii) accelerating (e.g. human activities altering natural slope stability) factors<sup>36</sup> along with gravitational forces. Most common one is shallow landslide, occurring in steep, soil mantled landscapes in different climatic zones.<sup>4,26,28,33</sup> This event has a significant role in an ability to change surface erosion patterns on a hill slope and sediment yields to nearby streams, triggered due to changes in topography, land cover and soil properties. The occurrence time and locations where landslides are likely to occur should thus be identified and modeled in advance in order to avoid or reduce the casualties. Casualties due to landslides caused by slope failures are significantly higher in the undulating terrains, resulting in huge economic losses due to higher value of endangered structures and the greater number of property and human casualties.

Landslides in mountainous terrain often occur during or after heavy rainfall, resulting in the loss of life and properties. Mapping or delineating areas susceptible to landslides is essential for moderating land use activities and implementing appropriate mitigation measures with

management options in mountainous areas.<sup>16</sup> Prediction of rainfall triggered hill slope disasters relied mostly on the valley slope<sup>8,20</sup>, rainfall intensity and duration that can cause hill slope failure.<sup>7,9</sup> During the last decade, theoretical models have been developed to predict landslide susceptibility based on topographic, geologic and hydrological variables as well as changes in watershed's land use using various parameters such as neural, fuzzy, object oriented, knowledge based etc.<sup>15,24,30,37,38,40,49,50,63</sup>

The availability of relatively detailed digital elevation data, coupled with simple slope-instability understanding and hill slope hydrological models, has led to refinements in physically-based modeling of shallow landslide hazards.<sup>22,62</sup> Daia and Leeb<sup>16</sup> illustrated slope instability modeling using Geographical Information Systems (GIS). Casadei et al<sup>12</sup> predict the time and location of landslides through a landslide warning system using a slope-instability analysis and hydrological models, which was validated by using historical data of landslide events for the period 1953–1998<sup>29</sup>.

The failure depth prediction model<sup>13</sup> is based on geotechnical properties (cohesion, angle of internal friction and density) of the soil and the local slope topography. Combination of failures and the consequent transportation of the failed materials lead to the evolution of a new hill-slope. These changes are due to the slope gradient, depending on the location of landslides, quantity and their run-out distances. There are numerous studies pertaining to shallow landslides across the globe<sup>41,58,59</sup> and on landslide hazard zonation and modeling<sup>3,24,51,52</sup>.

Most of the hazard appraisal techniques make use of Geographical Information Systems (GIS) along with temporal Remote Sensing data (RS) considering the complex spatial dynamics of the landslide process. Digital elevation models (DEMs) coupled with precipitation data at higher spatial and temporal resolution<sup>36</sup> aid in the shallow landslide hazard modeling.<sup>5,6,38,48,49,54,62</sup> Methods based on the infinite slope stability parameters<sup>31</sup> along with hydrological models<sup>37,49,62</sup>, have been used for the estimation of soil wetness spatially.<sup>2,11,18,37</sup>

Modelling these dynamic processes with appropriate methods require incorporation of all mechanisms acting at different spatial and temporal scale levels. However there are difficulties in the context of models of longer term landscape such as work on rapid mass movements has

concentrated on stability analysis, so that forecasting of destinations for slide debris is very inexact, even for an individual slide<sup>28</sup>, which has been addressed through various deterministic, heuristic and statistical based models to assess landslide incidence potential<sup>1,10,17,43</sup>. Deterministic methods deal with the estimation of quantitative values of stability variables namely soil strength, depth below the terrain surface, soil layer thickness, slope angle and water pressure, etc. of the landslide region with homogeneous intrinsic properties<sup>57,60</sup> and are applicable at large scale over small areas. However, the drawback of these deterministic model techniques is high degree of simplification due to paucity of data and at times it is prohibitive to acquire the required data.

## Methods

Openmodeller, a free and opensource software was used to model landslide occurrence locations. The model utilizes major contributors for landslides like precipitation and six site factors including aspect, DEM, flow accumulation, flow direction, slope, land cover, compound topographic index and historical landslide occurrence points. Precipitation in the wettest month and precipitation in the wettest quarter of the year were considered separately to analyse the effect of rainfall on hill slope failure for generating scenarios to predict landslides.

**Genetic Algorithm for Rule-set Prediction (GARP):** This is based on genetic algorithms<sup>23,25</sup> and has been widely used for predicting biological species<sup>55,56</sup>. GARP develops predictive models consisting of a set of conditional rules in the form of 'if-then' statements that describe the 'niches' of landslide occurrence points.

GARP uses a set of point localities where the landslide is known to have occurred along with a set of geographic layers representing the environmental parameters that might limit the landslide existence. Genetic Algorithms (GAs) are suitable for solving complex optimization problems and for applications that require adaptive problem-solving strategies. It maps strings of numbers to each potential solution and then each string becomes a representation of individual locations. Then the most promising in its search is manipulated for improved solution<sup>45</sup>. The GARP model is composed of a set of rules developed through evolutionary refinement, testing and selecting rules on random subsets of training data sets. Application of a rule set is more complicated as the prediction system must choose appropriate rule among a number of applicable rules. GARP maximizes the significance and predictive accuracy of rules without over-fitting.

Significance is established through a  $\chi^2$  test on the difference in the probability of the predicted value before and after application of the rule. GARP uses envelope rules, GARP rules, atomic and logit rules. In envelope rule, the conjunction of ranges for all of the variables is a climatic envelope or profile, indicating geographical regions where

the climate is suitable for that entity, enclosing values for each parameter. A GARP rule is similar to an envelope rule, except that variables can be irrelevant. An irrelevant variable is one where points may fall within the whole range. An atomic rule is a conjunction of categories or single values of some variables. Logit rules are an adaptation of logistic regression models to rules. A logistic regression is a form of regression equation where the output is transformed into a probability.<sup>56</sup>

**Support Vector Machine (SVM):** SVM are supervised learning algorithms based on heuristic algorithms of statistical learning theory<sup>27</sup>. SVM map input vectors to a higher dimensional space with maximal separating hyper plane. Two parallel hyper planes are constructed on each side of the hyper plane that separates the data. The separating hyper plane maximizes the distance between the two parallel hyper planes. An assumption is made that the larger is the margin or distance between these parallel hyper planes, better will be the accuracy of the classifier. The model produced by support vector classification depends only on a subset of the training data, because the cost function for building the model does not take into account training points that lie beyond the margin.<sup>27</sup>

In order to classify n-dimensional data sets, n-1 dimensional hyper plane is produced with SVMs. Fig. 1 illustrates various hyper planes separating two classes of data and there is only one hyper plane that provides maximum margin between the two classes as given in fig. 2 indicating the optimum hyper plane<sup>27</sup> and the points that constrain the width of the margin are the support vectors. SVMs locate a hyper plane that maximizes the distance from the members of each class to the optimal hyper plane in the binary case.

If there is a training data set containing m number of samples represented by  $\{x_i, y_i\}$  where  $(i = 1, \dots, m)$   $x \in \mathbb{R}^N$ , N-dimensional space and  $y \in \{-1, +1\}$  class label and the optimum hyper plane maximizes the margin between the classes. The hyper plane is defined as  $(w \cdot x_i + b = 0)$  given in fig. 3, where x is a point lying on the hyper plane, parameter w determines the orientation of the hyper plane in space, b is the bias that the distance of hyper plane from the origin. A separating hyper plane can be defined for two classes for the linearly separable case as:

$$w \cdot x_i + b \geq +1 \text{ for all } y = +1 \quad (1)$$

$$w \cdot x_i + b \leq -1 \text{ for all } y = -1 \quad (2)$$

These inequalities can be combined into a single inequality:

$$y_i (w \cdot x_i + b) - 1 \geq 0 \quad (3)$$

The training data points defined by the functions  $w \cdot x_i + b = \pm 1$ , are the support vectors and lie on these two hyper planes, which are parallel to the optimum hyper plane<sup>34</sup>.

The classes are linearly separable, if a hyper plane exists and satisfies inequality constraints as in (3), with the margin between these planes is equal to  $2/\|w\|$ <sup>34</sup>. Thus, the distance to the closest point is  $2/\|w\|$  and the optimum separating hyper plane can be found by minimizing  $\|w\|^2$  under the constraint (3).

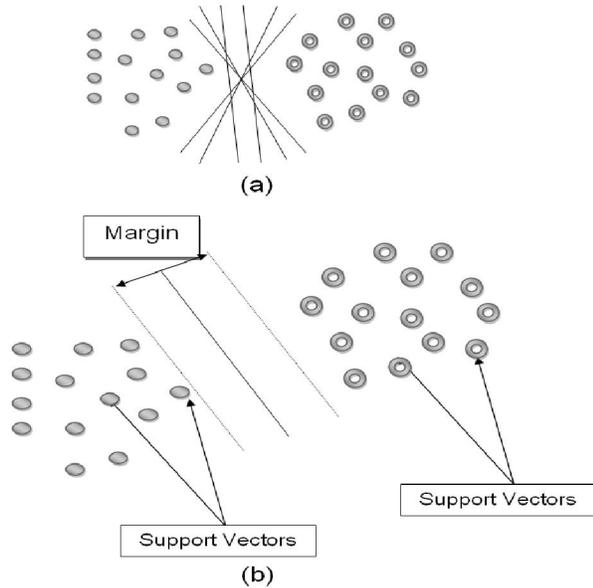


Fig.1: (a) Hyper planes for linearly separable data, (b) Optimum hyper plane and support vectors<sup>35</sup>

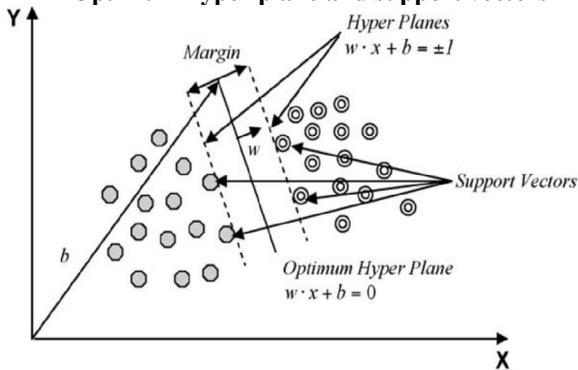


Fig. 2: Support vectors and optimum hyper plane for the binary case of linearly separable data sets<sup>35</sup>

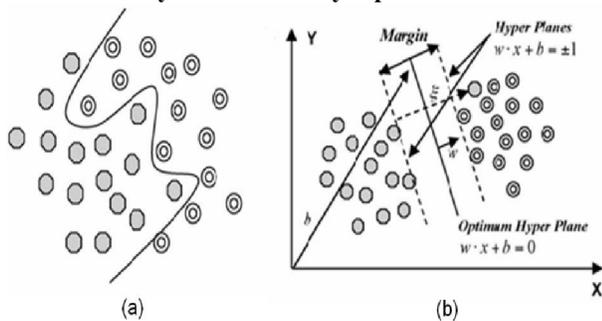


Fig. 3: (a) Separation of nonlinear data sets (b) Generalization of the solution by introducing slack variable for nonlinear data<sup>35</sup>

Determination of optimum hyper plane requires solving optimization function is given by:

$$\text{Min } (0.5 * \|w\|^2) \tag{4}$$

subject to the constraints,

$$y_i (w \cdot x_i + b) \geq -1 \text{ and } y_i \in \{+1, -1\} \tag{5}$$

Fig. 3 (a) illustrates that nonlinearly separable data is the case in various classification problems<sup>27</sup> as in the classification of remotely sensed images using pixel samples. In such cases, data sets cannot be classified into two classes with a linear function in input space. The technique can be extended to allow for nonlinear decision surfaces, when it is not possible to have a hyper plane defined by linear equations on the training data.<sup>14,44</sup> Considering this, the optimization problem is replaced by introducing  $\xi$  slack variable [Fig. 3(b)].

$$\text{min} \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^r \xi_i \right] \tag{6}$$

subject to constraints,

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, N \tag{7}$$

where C is a penalty parameter or regularization constant . This parameter allows striking a balance between the two competing criteria of margin maximization and error minimization, whereas the slack variables  $\xi_i$  indicates the distance of the incorrectly classified points from the optimal hyper plane. The larger is the C value, the higher is the penalty associated to misclassified samples.<sup>35</sup>

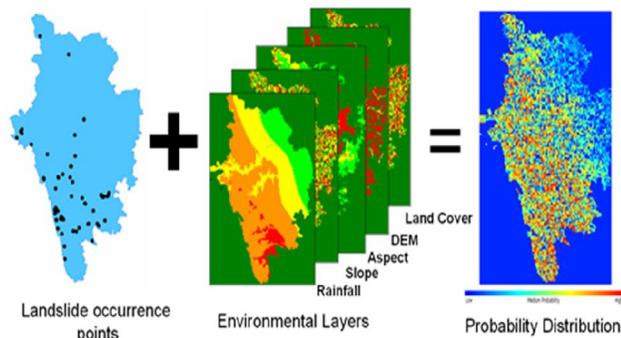
Data mapped into a higher dimensional space (H) through nonlinear mapping functions ( $\Phi$ ). An input data point x is represented as  $\Phi(x)$  in the high-dimensional space. The computation of  $[\Phi(x), \Phi(x_i)]$  is reduced by using a kernel function and the classification decision function would be:

$$f(x) = \text{sign} \left( \sum_i \alpha_i y_i K(x, x_i) + b \right) \tag{8}$$

where for each of r training cases there is a vector ( $x_i$ ) representing the respective spectral response together with a class membership ( $y_i$ ).  $\alpha_i$  ( $i = 1, \dots, r$ ) are Lagrange multipliers and  $K(x, x_i)$  is the kernel function. The magnitude of  $\alpha_i$  is determined by the parameter<sup>34</sup> and the kernel function enables the data points to spread in such a way that a linear hyper plane can be fitted<sup>19</sup>. The performance of SVMs varies depending on the choice of the kernel function and its parameters. Kernel functions used in SVMs are generally aggregated into four groups namely, linear, polynomial, radial basis function and sigmoid kernels.

OpenModeller<sup>39</sup> - a free and open source software was used for predicting the probable landslide areas. OpenModeller (<http://openmodeller.sourceforge.net/>) is a flexible, user

friendly, cross-platform environment and it includes facilities for reading landslide occurrence and environmental data, selection of environmental layers, creating a fundamental landslide prediction model and prediction with various environmental scenarios as shown in fig. 4.



**Fig. 4: Methodology used for landslide prediction in openModeller**



**Figure 5: Uttara Kannada district, Karnataka, India**

**Study area and data**

The Uttara Kannada district lies 74°9' to 75°10' east longitude and 13°55' to 15°31' north latitude, extending over an area of 10, 291 km<sup>2</sup> in the mid-western part of Karnataka state (Fig. 5). It accounts for 5.37 % of the total area of the state with a population above 1.2 million<sup>46</sup>. This region has gentle undulating hills, rising steeply from a narrow coastal strip bordering the Arabian Sea to a plateau at an altitude of 500 m with occasional hills rising above 600–860 m. This district with 11 taluks, can be broadly

categorised into three distinct regions — coastal lands (Karwar, Ankola, Kumta, Honnavar and Bhatkal taluks), mostly forested Sahyadrian interior (Supa, Yellapur, Sirsi and Siddapur taluks) and the eastern margin where the table land begins (Haliyal, Yellapur and Mundgod taluks). Climatic conditions range from arid to humid due to physiographic conditions ranging from plains, mountains to coast.

Survey of India (SOI) topo sheets of 1:50000 and 1:250000 scales were used to generate base layers – district and taluk boundaries, water bodies, drainage network etc. Field data were collected with a handheld GPS. Environmental data such as precipitation (rainfall) of wettest month and precipitation in the wettest quarter were downloaded from World Clim – Global Climate Data [<http://www.worldclim.org/bioclim>].

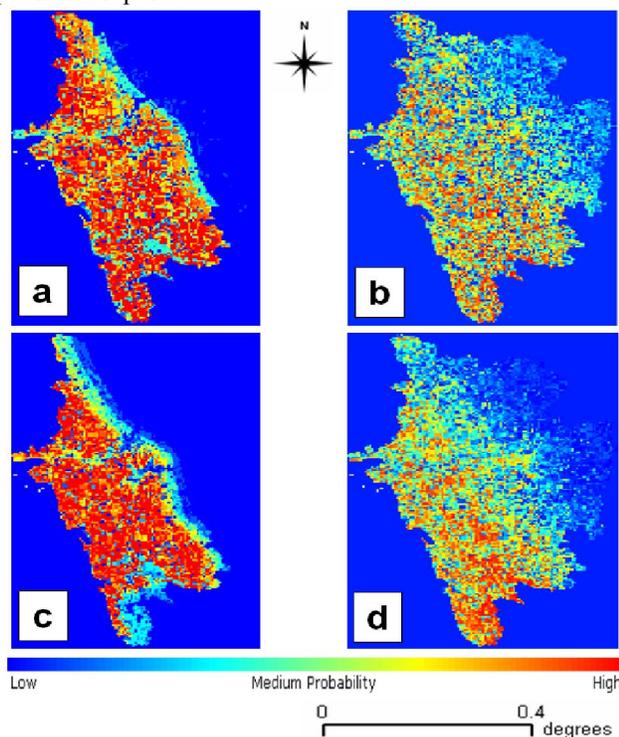
Other environmental layers used in the model were obtained from USGS Earth Resources Observation and Science (EROS) Center based Hydro1Kdatabase - [http://eros.usgs.gov/#/Find\\_Data/Products\\_and\\_Data\\_Available/topo30/hydro/asia](http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/topo30/hydro/asia)] which are as follows:

- i. Aspect describes the direction of maximum rate of change in the elevations or the slope direction. It is measured in positive integer degrees from 0 to 360, clockwise from north. Aspects of cells (pixels) of zero slope (flat areas) are assigned values of -1.
- ii. DEM is a digital representation of ground surface topography or terrain represented as a raster (a grid of squares) or as a triangular irregular network. DEMs are commonly built using remote sensing techniques or from land surveying.
- iii. Flow accumulation (FA) defines the number of pixels which flow into each downslope pixel. Since the pixel size of the HYDRO1k data set is 1 km, the flow accumulation value translates directly into upstream drainage areas in square kilometers. Values range from 0 at topographic highs to very large numbers (on the order of millions of square kilometers) at the mouths of large rivers.
- iv. Flow direction (FD) defines the direction of flow from each cell to its steepest down-slope neighbor derived from the hydrologically correct DEM. Values of flow direction vary from 1 to 255.
- v. Slope describes the maximum change in the elevations between each pixel and its eight neighbors expressed in integer degrees of slope between 0 and 90.
- vi. Compound Topographic Index (CTI) commonly referred to as the Wetness Index, is a function of the upstream contributing area and the slope of the landscape. It is calculated using the flow accumulation (FA) layer along with the slope as:

$$CTI = \ln\left(\frac{FA}{\tan(\text{slope})}\right) \quad (9)$$

In areas of no slope, a CTI value is obtained by substituting a slope of 0.001. This value is smaller than the smallest slope obtainable from a 1000 m data set with a 1 m vertical resolution.

The global land cover change maps were obtained from Global Land Cover Facility, Land Cover Change <http://glcf.umiacs.umd.edu/services/landcoverchange/landcover.shtml>; or <http://www.landcover.org/services/landcoverchange/landcover.shtml>. The spatial resolution of all the data were 1 km. The Google Earth data (<http://earth.google.com>) served in pre and post classification process and validation of the results. 125 landslide occurrence points of low, medium and high intensity were recorded using GPS from the field and published reports.



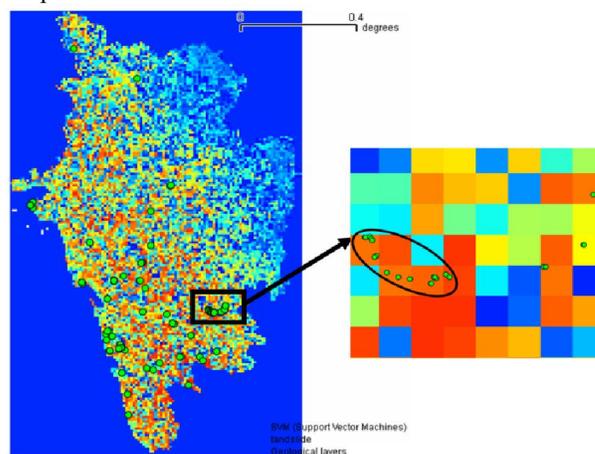
**Fig. 6: Probability distribution of the landslide prone areas**

**Results and Discussion**

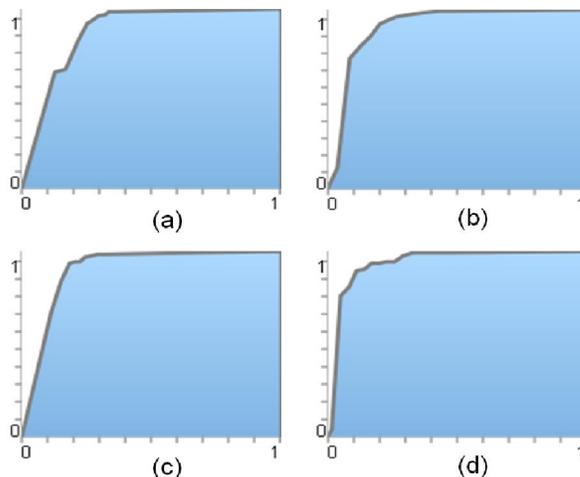
Precipitation of wettest month and wettest quarter of the year were used as different layers along with a set of seven other layers as mentioned earlier. Fig. 6 (a) and (b) indicate the landslide probability maps using GARP and SVM on precipitation of wettest month. The landslide occurrence points were overlaid on the probability maps to validate the prediction as shown in fig. 7. The GARP map had an accuracy of 92% and SVM map was 96% accurate with respect to the ground and Kappa values 0.8733 and 0.9083 respectively. The corresponding ROC curves are shown in fig. 7 (a) and (b). Total area under curve (AUC) for fig. 8

(a) is 0.87 and for fig. 8 (b) is 0.93. Figure 6 (c) and (d) are the landslide probability maps using GARP and SVM on precipitation of wettest quarter with accuracy of 91% and 94% and Kappa values of 0.9014 and 0.9387 respectively. ROC curves in figure 8 (c) and (d) show AUC as 0.90 and 0.94.

Various measures of accuracy were used as per Fielding and Bell to assess the outputs. Table I presents the confusion matrix structure indicating true positives, false positives, false negatives and true negatives. Confusion matrices were generated for each of the 4 outputs (Table II) and different measures of accuracy such as prevalence, global diagnostic power, correct classification rate, sensitivity, specificity omission and commission error computed and listed as in table III.



**Fig. 7: Validation of the probability distribution of the landslide prone areas by overlaying landslide occurrence points**



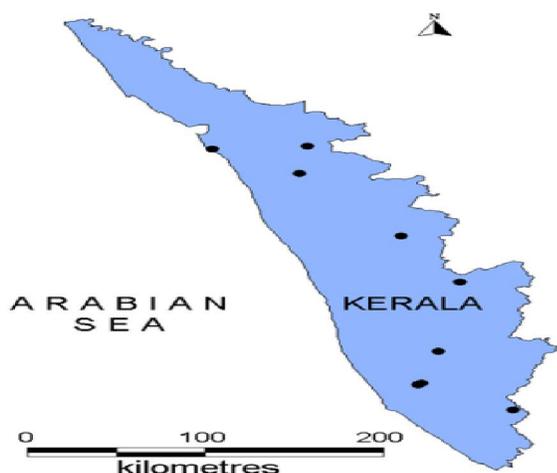
**Fig. 8: ROC curves for landslide prone maps (a) GARP (b) SVM on precipitation of wettest month; (c) GARP and (d) SVM on**

**Table 1**  
**Confusion Matrix**

	True Presence	True Absence
Predicted Presence	A	B
Predicted Absence	C	D

Key: A – True Positive, B – False Positive, C – False Negative, D – True Negative

The results indicate that the output obtained from SVM using precipitation of the wettest month was best among the 4 scenarios. It may be noted that the outputs from GARP for both the wettest precipitation month and quarter are close to the SVM in term of accuracy. One reason is that most of the areas have been predicted as probable landslide prone zones [indicated in red in fig. 6 (a) and (c)] with the highly undulating terrain consisting of steep slopes frequently exposed to landslides induced by rainfall. This was validated with the data collected from the field that showed overlap with the maximum number of landslide points occurring in the undulating terrain. This indicates that the region is more susceptible to landslides compared to north-eastern part of the district which has relatively flat terrain.

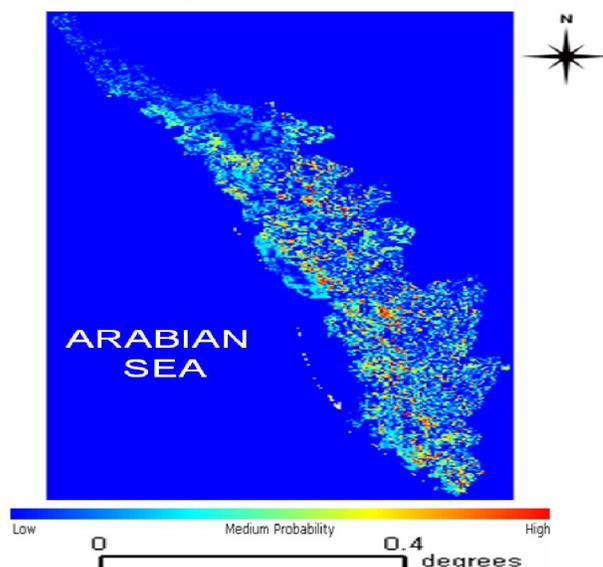


**Fig. 9: Landslide occurrence points in Kerala**

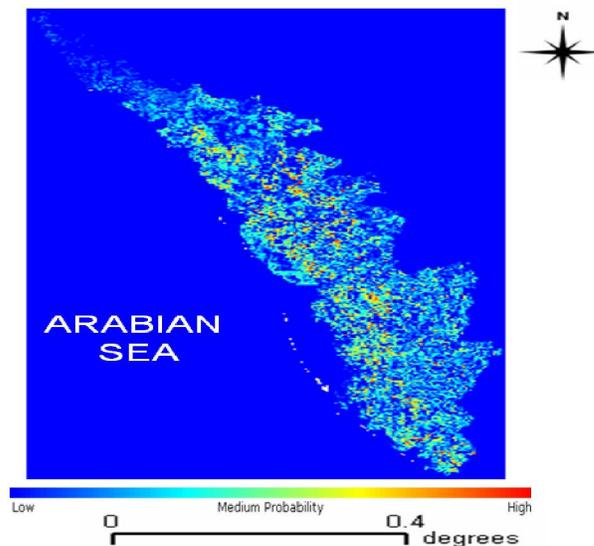
The exercise was repeated for another undulating terrain in the Western Ghats - Kerala state, India with the environmental layers along with 10 landslide occurrence field points as shown in fig. 9. The predicted output for precipitation in the wettest month is shown in fig. 10 and precipitation in the wettest quarter is shown in fig. 11 with an overall accuracy of 60%. The Kappa values for two cases were 0.966 and 0.965. The ROC curves are shown in fig. 12 (a) and (b) with 0.97 AUC for both the cases respectively. The confusion matrix and statistics are presented in table IV and V. The reason for low accuracy is due to lower field data collection (10 occurrence points)

throughout the state. Also, the coarse resolution of the pixel has contributed for poor accuracy.

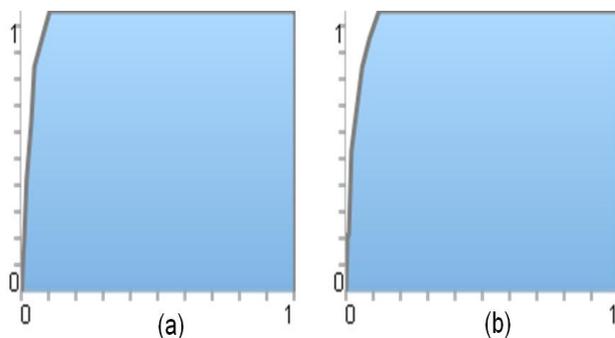
However, the Kappa value and the AUC are high which indicate that the probability distribution of the predicted points falls well within the occurrence points. The intensity of these landslide points may vary from low to medium to high. One potential limitation of the above data is their spatial resolution – 1 km. It is highly unlikely that landslides of this magnitude can occur in the study area. However, given the environmental layers along with the occurrence points, the probable areas of occurrence can be mapped with great certainty.



**Fig. 10: Probability distribution of the landslide prone areas in Kerala using precipitation in the wettest month data**



**Fig. 11: Probability distribution of the landslide prone areas in Kerala using precipitation in the wettest quarter data**



**Fig. 12: ROC curves for the predicted landslide prone maps of Kerala**

**(a) using precipitation of the wettest month, (b) using precipitation of the wettest quarter**

The movement of a slope is complex and is induced or perturbed by many other factors besides rainfall such as groundwater table, soil moisture etc. which were not used in the prediction. It is known that there are accelerations of slope movements after intense rainfall, however, to what extent rainfall affects the slope movement remains unknown and the correlation between rainfall and the slope movement could be ambiguous in absence of detailed observations. Timely rainfall, soil moisture, grain-size, lithology, geological structure, seismological observations and the longer time intervals for which data are available can improve the accuracy of the model.

Beyond their key role in identifying and mapping the landslides, the choice of the variables (environmental layers) used for the prediction is also closely involved. Results can be improved with additional supplementary data such as hydrological field data etc. However, the results obtained in the current study are significant as most influential variables of landslide were used.

Natural disasters have drastically increased over the last decades. National, state and local government including NGOs are concerned with the loss of human life and damage to property caused by natural disasters. The trend of increasing incidences of landslides occurrence is expected to continue in the next decades due to unplanned anthropogenic activities leading to large scale land cover changes, deforestation and increased regional precipitation in landslide-prone areas due to changing climatic patterns<sup>47</sup>. Modern technologies with prediction ability help to mitigate the effects of natural hazards including landslides. In fact, to ensure hazard free in mountainous regions, the development of urban facilities such as homes and new roads must consider geological and geotechnical factors. Therefore, it is desired to have a notion about the main factors controlling the slope instability, assessing its severity, discriminating areas with presence/absence of landslides, updating and interpreting landslide data and determining areas that are prone to landslides.

## Conclusion

Landslides occur when masses of rock, earth or debris move down a slope. Mudslides, debris flows or mudflows, are common type of fast-moving landslides that tend to flow in channels. These are caused by disturbances in the natural stability of a slope which is triggered by high intensity rains. Mudslides usually begin on steep slopes and develop when water rapidly collects in the ground and results in a surge of water-soaked rock, earth and debris. Causal factors may be either preparatory or triggering. Preparatory causes are factors which have made the slope potentially unstable. The triggering cause is the single event that finally initiated the landslide. Thus, causes combine to make a slope vulnerable to failure and the trigger finally initiates the movement. Thus a landslide is a complex dynamic system. This characteristically involves many different processes operating together, often with differing intensity during successive years.

The primary criteria that influence landslides are precipitation intensity, slope, soil type, elevation, vegetation and temporal changes in land cover. The present study demonstrated the effectiveness of two pattern recognition techniques: Genetic Algorithm for Rule-set Prediction and Support Vector Machine. The landslide hazard prediction study conducted in Uttara Kannada and Kerala has shown that these techniques with small datasets can yield landslide susceptibility maps of significant predictive power. The efficiency of the model has been demonstrated by the successful validation. However, when the predicted features may have different immediate causes, one should carefully avoid including triggering factors among the predictor variables since they restrict the scope of the prediction map and convey often a poorly constrained time dimensions.

The reliability of the susceptibility map fundamentally depends on the quality of the data and sample size apart from appropriately validation. The analysis showed that SVM applied on precipitation data of the wettest month with 96% accuracy was close to reality for Uttara Kannada district and GARP applied on precipitation data of the wettest quarter was more successful in identifying the landslide prone areas in Kerala.

## Acknowledgement

We are grateful to ISRO-IISc Space Technology Cell, Indian Institute of Science and NRDMS division, Ministry of Science and Technology, DST, Government of India for the financial support. The environmental layers were obtained from World Clim - Global Climate Data. NRSA, Hyderabad provided the LISS IV data used for land cover analysis. We thank USGS Earth Resources Observation and Science (EROS) Center for providing the environmental layers and Global Land Cover Facility (GLCF) for facilitating the Land Cover Change product.

**Table 2**  
**Confusion matrix for GARP and SVM Outputs for Uttara Kannada**

Uttara kannada		True presence	True absence	Number of usable presence	Number of usable absence
GARP WITH PRECIPITATION OF WETTEST MONTH	Predicted Presence	120	0	125	0
	Predicted Absence	5	0		
SVM WITH PRECIPITATION OF WETTEST MONTH	Predicted Presence	118	0	125	0
	Predicted Absence	7	0		
GARP WITH PRECIPITATION OF WETTEST QUARTER	Predicted Presence	118	0	125	0
	Predicted Absence	7	0		
SVM WITH PRECIPITATION OF WETTEST QUARTER	Predicted Presence	117	0	125	0
	Predicted Absence	8	0		

**Table 3**  
**Statistics of GARP and SVM Outputs for Uttara Kannada**

UTTARA KANNADA	PREVALENCE (A+C)/N	GLOBAL DIAGNOSTIC POWER (B+D)/N	CORRECT CLASSIFICATION RATE (A+D)/N	SENSITIVITY A/(A+C)	SPECIFICITY D/(B+D)	OMISSION ERROR C/(A+C)	COMMISSION ERROR B/(B+D)
GARP WITH PRECIPITATION OF WETTEST MONTH	-	-	0.96	0.96	-	0.04	-
SVM WITH PRECIPITATION OF WETTEST MONTH	-	-	0.94	0.94	-	0.06	-
GARP WITH PRECIPITATION OF WETTEST QUARTER	-	-	0.94	0.94	-	0.06	-
SVM WITH PRECIPITATION OF WETTEST QUARTER	-	-	0.94	0.94	-	0.06	-

\* Key: A – True Positive, B – False Positive, C – False Negative, D – True Negative, N – Number of Samples.

**Table 4**  
**Confusion matrix for GARP Outputs for Kerala**

Kerala		True presence	True absence	Number of usable presence	Number of usable absence
GARP with precipitation of wettest month	Predicted Presence	6	0	10	0
	Predicted Absence	4	0		
GARP with precipitation of wettest Quarter	Predicted Presence	6	0	10	0
	Predicted Absence	4	0		

**Table 5**

## STATISTICS of GARP Outputs for Kerala

Kerala	Prevalence (A+C)/N	Global diagnostic Power (B+D)/N	Correct Classification Rate (A+D)/N	Sensitivity A/(A+C)	Specificity D/(B+D)	Omission Error C/(A+C)	Commision Error B/(B+D)
GARP with precipitation of wettest month	-	-	0.6	0.6	-	0.4	-
GARP with precipitation of wettest Quarter	-	-	0.6	0.6	-	0.4	-

\* Key: A – True Positive, B – False Positive, C – False Negative, D – True Negative, N – Number of Samples.

## References

- Atkinson P.M. and Massari R., Generalised linear modelling of susceptibility to landsliding in the central Appennines, Italy, *Computer and Geosciences*, **24**(4), 371–383 (1998)
- Barling R.D., Moore I.D. and Grayson R.B., A quasi-dynamic wetness index for characterizing the spatial distribution of zones of surface saturation and soil water content, *Water Resources Research*, **30**, 1029–1044 (1994)
- Bathurst J.C., Burton A., Clarke B.G. and Gallart F., Application of the SHETRAN basin-scale, landslide sediment yield model to the Llobregat basin, Spanish Pyrenes, *Hydrological Processes*, **20**, 3119–3138 (2006)
- Benda L.E. and Cundy T.W., Predicting deposition of debris flows in mountain channels, *Canadian Geotechnical Journal*, **27**, 409–417 (1990)
- Borga M., Dalla G., Fontana, Da Ros D. and Marchi L., Shallow landslide hazard assessment using a physically based model and digital elevation data, *Environmental Geology*, **35**, 81–88 (1998)
- Burton A. and Bathurst J.C., Physically based modelling of shallow landslide sediment yield at a catchment scale, *Environmental Geology*, **35**, 89–99 (1998)
- Caine N., The rainfall intensity–duration control of shallow landslides and debris flows, *Geografiska Annaler*, **62**, 23–27 (1980)
- Campbell R.H., Soil slips, debris flow and rainstorms in the Santa Monica Mountain and vicinity Southern California, US Geological Survey Professional Paper, 851 (1975)
- Cannon S.H. and Ellen S., Rainfall conditions for abundant debris avalanches San Francisco Bay Region, California, *California Geology*, **38**, 267–272 (1985)
- Carrara A., Multivariate models for landslide hazard evaluation, *Mathematical Geology*, **15**(3), 403–426 (1983)
- Casadei M., Dietrich W.E. and Miller N.L., Testing a model for predicting the time and location of shallow landslide initiation in soil-mantled landscapes, *Earth Surface Processed and Landforms*, **28**, 925–950 (2003)
- Junjie Sun, Lanmin Wang and Guangqi Chen, A Probabilistic Method to assess the Regional Susceptibility of Landslides induced by Earthquake in Kitakyushu City, Japan, *Disaster Advances*, **4**(1), 7-18 (2011)
- Claessens L., Schoorla J.M. and Veldkampa A., Modelling the location of shallow landslides and their effects on landscape dynamics in large watersheds: An application for Northern New Zealand, *Geomorphology*, **87**(1-2), 16-27 (2007)
- Cortes C. and Vapnik V., Support-vector network, *Machine Learning*, 273–297 (1995)
- Crosta G.B. and Frattini P., Rainfall-induced Landslides and debris flows, *Hydrological Processes*, **22**, 473-477 (2008)
- Daia F.C. and Leeb C.F., Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, *Geomorphology*, **42**(3-4), 213-228 (2002)
- DeGraff J. and Romesburg H., Regional landslide-susceptibility assessment for wildland management: a matrix approach In Coates D., Vitek J., eds., George Allen and Unwin, Thresholds in Geomorphology, London, 401–414 (1980)
- Dietrich W.E., Reiss R., Hsu M. and Montgomery D.R., A process-based model for colluvial soil depth and shallow landsliding using digital elevation data, *Hydrological Processes*, **9**, 383–400 (1995)
- Dixon B. and Candade N., Multispectral land use classification using neural networks and support vectormachines: one or the other, or both?, *International Journal of Remote Sensing*, **29**, 1185–1206 (2008)
- Ellen S.D. and Wiczorek G.K., Landslides, floods and marine effects of the storm of January 3-5, 1982, in the San Francisco Bay Region, California, United States Geological Survey Profession Paper, 1434 (1988)
- Fielding A.H. and Bell J.F., A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental Conservation*, **24**, 38-49 (1997)
- Goetza J.N., Guthrieb R.H. and Brenninga A., Integrating physical and empirical landslide susceptibility models using generalized additive models, *Geomorphology*, **129**(3-4), 376-386 (2011)

23. Goldberg D.E., Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley (1989)
24. Guzzetti F., Carrara A., Cardinali M. and Reichenbach P., Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, *Geomorphology*, **31**, 181–216 (1999)
25. Holland J.H., Adaptation in Natural and Artificial Systems, An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence, University of Michigan Press, Ann Arbor (1975)
26. Hutchinson J.N., Morphological and geotechnical parameters of landslides in relation to geology and hydrology, In Proceedings 5th International Symposia on Landslides, Lausanne, 3–36, (1988)
27. Kavzoglu T. and Colkesen I., A kernel functions analysis for support vector machines for land cover classification, *International Journal of Applied Earth Observation and Geoinformation*, **11**, 352–359 (2009)
28. Kirkby M.J., General models of long-term slope evolution through mass movement In Anderson M.G. and Richards K.S., Editors, Slope Stability, Wiley, Chichester, 359–379 (1987)
29. Kwan T.L. and Jui Yi H., Prediction of landslide occurrence based on slope-instability analysis and hydrological model simulation, *Journal of Hydrology*, **375**, 489–497 (2009)
30. Kwang Yeona Y., Hana J. and Ryub K., Landslide susceptibility mapping in Injae, Korea, using a decision tree, *Engineering Geology*, **116(3-4)**, 274–283 (2010)
31. Lambe W.R. and Whitman R.V., Soil Mechanics, 2nd ed., Wiley, New York (1979)
32. Mathur A. and Foody G.M., A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1335–1343 (2004)
33. Mathur A. and Foody G.M., Multiclass and binary SVM classification: Implications for training and classification users, *IEEE Geoscience and Remote Sensing Letters*, **5**, 241–245 (2008a)
34. Mathur A. and Foody G.M., Crop classification by support vector machine with intelligently selected training data for an operational application, *International Journal of Remote Sensing*, **29**, 2227–2240 (2008b)
35. Melgani F. and Bruzzone L., Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1778–1790 (2004)
36. Monia S., Salvatore G., Fernando N., Andrea P. and Maria R. C., Pre-processing algorithms and landslide modelling on remotely sensed DEMs, *Geomorphology*, **113**, 110–125 (2009)
37. Montgomery D.R. and Dietrich W.E., A physically based model for the topographic control on shallow landsliding, *Water Resources Research*, **30(4)**, 1153–1171 (1994)
38. Montgomery D.R., Sullivan K. and Greenberg H.M., Regional test of a model for shallow landsliding, *Hydrological Processes*, **12**, 943–955 (1998)
39. Munoz M.E.S., Giovanni R., Siqueira M.F., Sutton T., Brewer P., Pereira R.S., Canhos Dal and Canhos V.P., OpenModeller- a generic approach to species' potential distribution modeling, *Geoinformatica* (2007)
40. Nandia A. and Shakoorb A., A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses, *Engineering Geology*, **110(1-2)**, 11–20 (2009)
41. Okura Kitahara H., Ochiai H., Sammori T. and Kawanami A., Landslide fluidization process by flume experiments, *Engineering Geology*, **66**, 65–78 (2002)
42. Oommen T., Misra D., Twarakavi N.K.C., Prakash A., Sahoo B. and Bandopadhyay S., An objective analysis of support vector machine based classification for remote sensing, *Mathematical Geosciences*, **40**, 409–424 (2008)
43. Pachauri A.K. and Pant M., Landslide hazard mapping based on geological attributes, *Engineering Geology*, **32(1-2)**, 81–100 (1992)
44. Pal M. and Mather P.M., Support vector machines for classification in remote Sensing, *International Journal of Remote Sensing*, **26**, 1007–1011 (2003)
45. Pujari A.K., Data Mining Techniques, University Press, Hyderabad (2006)
46. Ramachandra T.V. and Nagarathna A.V., Energetics in paddy cultivation in Uttara Kannada district, *Energy Conservation and Management*, **42(2)**, 131–155 (2001)
47. Ramachandra T.V., Subashchandran M.D. and Anant Hegde Ashisar, Landslides at Karwar, October 2009: Causes and Remedial Measures, ENVIS Technical Report: 32, Environmental Information System, Centre for Ecological Sciences, Bangalore, (2009)
48. Romeo R., Seismically induced landslide displacements: a predictive model, *Engineering Geology*, **58**, 337–351 (2000)
49. Rosso R., Rulli M.C. and Vannucchi G., A physically based model for the hydrologic control on shallow landsliding, *Water Resources Research*, **42**, W06410 (2006)
50. Saitoa H., Nakayamaa D. and Matsuyamaa H., Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence, The Akaishi Mountains, Japan, *Geomorphology*, **109(3-4)**, 108–121 (2009)
51. Schuerch P., Densmore A.L., Mc Ardell B.W. and Molnar P., The influence of landsliding on sediment supply and channel change in a steep mountain catchment, *Geomorphology*, **78**, 222–235 (2006)
52. Schwab M., Rieke-Zapp D., Schneider H., Liniger M. and Schlunegger F., Landsliding and sediment flux in the Central

- Swiss Alps: a photogrammetric study of the Schimbrig landslide, Entlebuch, *Geomorphology*, **97**, 392–406 (2008)
53. Selby, Hillslope Materials and Processes, Oxford University Press, Oxford, (1993)
54. Shou K.J. and Wang C.F., Analysis of the Chiufengershan landslide triggered by the 1999 Chi-Chi earthquake in Taiwan, *Engineering Geology*, **68**, 237–250 (2003)
55. Stockwel D.R.B., Evaluation and Training in the Use and Interpretation of GARP Genetic Algorithms for Prediction, Consultancy Report to Australian Nature Conservation Agency (1995)
56. Stockwell D.R.B. and Peters D.P., The GARP modeling system; Problems and solutions to automated spatial prediction, *International Journal of Geographic Information Systems*, **13**, 143-158 (1999)
57. Turner K. and Jayaprakash G., Introduction, In Turner A.K., Schuster R.L., eds., Landslides: Investigation and Mitigation, Special Report, National Academic Press, Washington, DC, **247**, 3– 12 (1992)
58. Wang G. and Sassa K., Factors affecting rainfall-induced flowslides in laboratory flume tests, *Geotechnique*, **51**, 587–599 (2001)
59. Wang G. and Sassa K., Pore-pressure generation and movement of rainfall-induced landslides: effects of grain size and fine-particle content, *Engineering Geology*, **69**, 109–125 (2003)
60. Ward T., Ruh-Ming L. and Simons D., Mapping landslide hazard in forest watershed, *Journal of Geotechnical Engineering Division*, **108(GT2)**, 319– 324 (1988)
61. Wiczorek G.F., Effect of rainfall intensity and duration on the debris flows in central Santa Cruz Mountains, California, *Geological Society of America Reviews in Engineering Geology*, **7**, 93–104 (1987)
62. Wu W. and Sidle R., A distributed slope stability model for steep forested basins, *Water Resources Research*, **31(8)**, 2097–2110 (1995)
63. Yao X., Thamb L.G. and Daia F.C., Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China, *Geomorphology*, **101(4)**, 572-582 (2008).

(Received 09<sup>th</sup> October 2012, accepted 20<sup>th</sup> December 2012)