International Journal of Soft Computing 13 (2): 31-37, 2018
ISSN: 1816-9503

Soft Bayesian Model for Landslide Risk Analysis

J. Velmurugan and M. Venkatesan
School of Computing Science and Engineering, VIT University, Vellore, India

Abstract: A natural disaster causes huge loss in terms of people life and infrastructures. Landslide is one of the prime disasters in the hill regions such as Uttarakhand, Sikkim and Ooty in India. The extent of damages of landslide could be reduced or minimized by proposing novel landslide risk analysis model. Landslide is generated by various factors such as rainfall, soil, slope, land use and land covers, geology, etc. Data science and soft computing plays major role in landslide risk analysis. In this study, classification data science technique is integrated with rough set model and Soft Bayesian Prediction Model (SBPM) is proposed to analyze the possibilities of various landslide risk level at Coonor Taluk of Niligiri District. The proposed model is validated with real time data and performance is compared with other classification models.

Key words: GIS, rough set, Bayesian, landslide, disaster, real

INTRODUCTION

Environmental disasters like cyclone, earthquakes, rainfall, tsunamis and landslides cause incalculable deaths and fearful damage in the world and the changes in environment infrastructure. Landslides are the main disaster which causes huge damages in the infrastructure and incalculable of deaths are happening in the every year. Landslide predictions are to be identified using different statistical methods based on the Geographical Information System (GIS) technology. In the past few year, many research are carried out to identify the landslides in order to improve the efficiency of landslide prediction. In these cases few research may success in the prediction of landslide using data mining technologies. In this study, we consider the Coonoor District, Tamilnadu because the huge landslides are happening in the every year due to heavy rainfall. As number of slidings and floodings have been happening in many regions around the world in the recent year. In this connection, we have to improve our ability to deal with the natural hazards and risks.

There are so any factors are involving in the landslides but the effects of the landslides are different area to area. The mathematical bond between the factors which impact landslide and the landslide strength prediction is solid to obtain. However, it is a comparatively exact method to get a mathematical investigation model with the historical data. To get a reliable prediction result of landslides we have to evaluate the geological and environmental circumstances and consider positive factors such as the soil, slope, rainfall, land use and land cover, geomorphology and geology.

In the existing studies, statistical models, neural networks, fuzzy based neural networks and data mining classification techniques are applied to solve these problems. To improve the ability of the existing research we are proposed a new model called Soft Bayesian Prediction Model (SBPM) to predict the landslides which may give accurate results. There are so many components are involved to identify the landslides. In these few, components may fully depend for the prediction of landslides. The importance of the attributes is identified by using the soft computing approach is rough set theory. To get an exact and consistent prediction result of landslides we consider the few of environmental factors such as the soil, slope, rainfall, land use and land cover, geomorphology and geology. In this study, the above said factors are considered to construct a novel Soft Bayesian Prediction Model (SBPM) for the analysis of risk level at Coonoor Taluk of Nilgiri District.

Literature review: Landslides are occurred due to heavy rainfall, earthquake ground motion and other relative environmental factors. In addition to that the environmental factors are also, considered for the landslide occurrence such as soil, slope, land use and land cover, slope geology and morphological parameters. The existing studies had focus on mathematical methods including analytical hierarchy method, data mining approaches, soft computing techniques. Likelihood ratio and artificial neural networks (Chung, 2006; Melchiorre *et al*., 2008; Nefeslioglu *et al*., 2008; Wu and Chen, 2009). However, the accuracy of these methods is calculated by manually using mathematical methods. The disadvantages of these methods are the

Corresponding Author: J. Velmurugan, School of Computing Science and Engineering, VIT University, Vellore, India

value and relative weightings are manually assigned the professed influence on the occurrence of landslides. In additional to that these methods are integrated with ecological factors. In authenticity, only a few factors or combinations of factors had an enormous input to identify the landslides.

To differentiate these factors are very difficult so that landslide susceptibility can be exactly mapped. The present research are carried out in the data mining area has created more awareness in the knowledge discovery for landslide susceptibility prediction (Gorsevski and Jankowski, 2008; Wan, 2009; Wan *et al*., 2010). For example, the environmental factors are categorized and evaluated systematically by using data mining classification method, i.e., decision tree (an entropy based) to constructing the knowledge rules. Naturally, knowledge for landslide forecast can be supposed as a body of data which constitutes our domain of interest. The knowledges are represented in the form of table, columns of which are labeled by attributes and rows by environmental conditional factor values are presence. The decision tree problems for the selection of environmental factors can be formulated using the above decision Table formalism to predict the landslide susceptibility. In the present study, the rough set theory developed by Pawlak (1982) was considered for reduction of a knowledge Table in such a manner that probability can be considered with a less number of attributes. Also in the rough set processing, the landslide susceptibility database are regarded as a decision knowledge Table which consist of condition attributes (factors affecting landslides) and decision attributes (landslide presence or no landslide presence).

The rough set theory dealing with completely different from the traditional mathematical analyses that understand distributions in the independent variables (Arciszewski and Ziarko, 1990; Beynon, 2001; Pawlak and Slowinski,1994; Anbalagan and Chandrasekaran, 2015; Pawlak, 1997; Slowinski *et al*., 2012; Wan, 2009; Saxena *et al*., 2014; Yilmaz, 2009; Zeng *et al*., 2006; Zhang *et al*., 2000; Rouse Jr. *et al*., 1973). The Bayesian classification rough sets are invented for making classification decisions based on available knowledge information. Bayesian authentication rough sets are proposed for weighting pieces of evidence given by correspondence classes. The models can be studied with respect to three basic issues. For calculating the thresholds we have a systematic method for a Bayesian classification rough set model according to Bayesian decision theory (Yao and Zhou, 2016). Rough set theory is one of the mathematical model that provides a various statistical concept to extract the feature knowledge from real data involving ambiguity, uncertainty and impreciseness and the extracted knowledge will be supplied successfully in the field of machine learning, pattern recognition and knowledge discovery (Saito *et al*., 2009). The analyzed landslide risk by weighted decision tree prediction model. The 4 important landslide induced factors such as rainfall, land use/land cover, slope and geology are considered for the analysis. The study contains remote sensing images and field data are used to prepare various thematic maps. The performance of the weighted decision tree prediction model is compared with existing classification approaches. Weighted decision tree prediction model is more suitable and accurate than decision tree classifier (Anbalagan and Chandrasekaran, 2015). The weighted decision tree prediction modeling approach, combined with the use of remote sensing and Geographical Information Systems (GIS) spatial data, yields a reasonable accuracy in the landslide risk analysis.

MATERIALS AND METHODS

Soft Bayesian Prediction Model (SBPM)

Rough set approach: Rough set theory was developed by the Zdzislaw (Pawlak, 1982). It has created the awareness too many research scholars all over the countries which contributed fundamentally to its development and applications.

It also found many attractive applications. It seems that the rough set approaches are involved in the fundamental importance of artificial intelligence, soft computing and cognitive sciences, especially in the areas of machine learning, knowledge mining, decision analysis and pattern recognition.

The major benefits of rough set theory in data analysis are that it does not require any preliminary or additional information about it. The proposed approach of rough set theory:

- Provides competent algorithms to find hidden patterns in knowledge
- Data reduction
- Evaluates significance values of data
- Create the sets of decision rules from given knowledge
- Easy to understand
- Offers clear-cut explanation of obtained results
- Most algorithms based on the rough set theory are particularly appropriate for parallel processing

Definition 1: RS = (A, X, V, f) are set to an information system. Among them A = $\{A_1, A_2, A_3, ..., A_n\}$ is non empty finite sets which is called the domain space, X = $\{x_1, x_2, ..., x_n\}$ is non-empty finite attribute set which is called the attribute set, V = $\cup X_a$, x ∈ X , V_x is attribute's domain range, f: A X X->Vx is the information function. When x is a x has unique value in V_a. On the side for sequence C ($c_1(x)$, $C_2(x)$, ..., $C_n(x)$) and sequence D (d_1(x) d_2 (x), ..., d_n (x)) B = C∩D, C∪D = Ø , S = (A, X,V, f)

is called as decision table of the information system. $C_1(x)$, $C_2(x)$, ..., $C_n(x)$ is called as the condition attribute set.

Definition 2: For the given knowledge representation system S = (A, X, V , f) the in-discernable relationship of any attribute is as follows:

$$RIND\ (R) = : \{(x, y) \in AX\ A: \forall a \in B(f(x,a) = f(y,a)\}$$

Definition 3: For the given knowledge representation system S = (A, X, V, f), P⊆X, V⊆, x∈A the lower and upper approximation set for A with regard to RIND B is as respectively:

$$R(B) = \cup \{a \in A: RIND\ (X) \subseteq B\}$$
$$R(B) = \cup \{a \in A: RIND\ (X) \cap B = \emptyset\}$$

Definition 4: For the given knowledge representation system S = (A, X,V, f) if, P, Q⊆X , the positive domain POS p(Z) is defined as:

$$POSp\ (Z) = \cup R(A)$$
$$x \subseteq U/P$$

Among them, R(A) is the lower approximation of A. Let P, Z⊆C∪D, the given the upper and lower approximations P(B) and P(B) the P-positive region of A can be defined as:

$$POSp(Z) = \cup \{\ \underline{PB}: B \in A/RIND(Z)\}$$

The positive region $POS_P(Z)$ contains all the object sin A that can be classified into one class without an error defined by RIND(Z). The bound region can be defined as PB-PB and the negative region as A-PB. The dependency of Z on P is defined as:

$$Yp(Z) = \frac{(Card\ (POS_P(Z))}{(card(A))}$$

A measure of significance of the attribute a∈P from the set P with respect to the classification A/RIND(Z) generated by a set Q is:

$$\mu p, Z(a) = \frac{(Card\ (POS_P(Z))\text{-}card\ (POS_p\ (a)(Z))}{(card(A))}$$

A measure of the accuracy of an approximation of a set in the space P is defined as:

$$\mu(B) = (R(B))/(\overline{R(B)})$$

Naive bayes classifier: The data mining technique have various of classification algorithms like Support Vector Machine (SVM) decision tree, back propagation neural network, rule based classification and Naive Bayesian classification. Naive Bayesian is the technique is classifying data based on the decision attribute. A Naive Bayesian classifier method is a set of a supervised learning algorithm based on applying Baye's theorem with the "Naive" assumption of liberty between each pair of features. Given a class variable Y and a dependent feature vector x_1 through x_n, Baye's theorem states the following relationship:

$$P(Y \mid x_1,...,x_n)\text{-}\frac{P(y)P(x_1,...,x_n)}{P(x_1,...,x_n)}$$

Where:

P(y|x) = The posterior probability of decision attribute class (C, target) given predictor (X, attributes)
P(y) = The prior probability of C
P(x|y) = The likelihood which is the probability of predictor given class
P(x) = The prior probability of predictor

There are m layers of spatial map data containing "causal" factors which are known to associate with the occurrences of future landslides in the study area. The significant landslide factors from the study area are extracted from the map sources and represented in Table 1. Assume that, we have 5 classes of landslide susceptibility C1 = High, C2 = Low, C3 = Medium, C4 = Very high, C5 = Very low.

Let A be a training set of tuples and their connected with class labels. Each tuple is mentioned by an n dimensional attribute vector, $X = (x_1, x_2, \ldots, x_n)$ depicting m measurements made on the tuple from j attributes, respectively $A_1, A_2, \ldots, A_j$. Suppose that there are i classes C1, C2, …, Ci. Given using the naive independence assumption that:

$$P(x_i|y, x_1, ..., x_{i-1}, X_{i+1}, ..., x_n) = P(x_i|y)$$

for all i this relationship is simplified to:

$$P(y \mid x_1,...,x_n)\ \frac{p(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1,...,x_n)}$$

Since, $P(x_1, \ldots, x_n)$ is constant given the input:

Table 1: Training landslide data set

FID	Soil	Slope	Rainfall	Land use land cover	Geomorphology	Geology	Zones
1	Clayey	0-8.029984	109.397533-124.594088	Agriculture	Ridge type structural hills (Large)	Charnockite group	Moderate
2	Clayey	0-8.029984	135.63122-150.827775	Scrub forest	Ridge type structural hills (large)	Charnockite group	Moderate
3	Clayey	0-8.029984	135.63122-150.827775	Scrub forest	Ridge type structural hills (large)	Charnockite group	Moderate
4	Clayey	0-8.029984	150.827775-171.75126	Forest plantations	Ridge type structural hills (large)	Charnockite group	High
5	Clayey	0-8.029984	150.827775-171.75126	Land without scrub	Ridge type structural hills (large)	Charnockite group	High
6	Clayey	0-8.029984	150.827775-171.75126	Scrub forest	Ridge type structural hills (large)	Charnockite group	High
7	Clayey	0-8.029984	150.827775-171.75126	Scrub forest	Ridge type structural hills (large)	Charnockite group	Low
8	Clayey	0-8.029984	150.827775-171.75126	Scrub forest	Ridge type structural hills (large)	Charnockite group	Low
9	Clayey	0-8.029984	171.75126-200.559911	Forest plantations	Ridge type structural hills (large)	Charnockite group	High
10	Clayey	0-8.029984	59.665397-88.474048	Dense forest	Ridge type structural hills (large)	Gneiss	Low
11	Clayey	0-8.029984	59.665397-88.474048	Dense forest	Ridge type structural hills (large)	Gneiss	Very low
12	Clayey	0-8.029984	59.665397-88.474048	Land with scrub	Ridge type structural hills (large)	Gneiss	Low
13	Clayey	0-8.029984	59.665397-88.474048	Land with scrub	Ridge type structural hills (large)	Gneiss	Moderate
14	Clayey	0-8.029984	59.665397-88.474048	Scrub forest	Ridge type structural hills (large)	Gneiss	Low
15	Clayey	0-8.029984	59.665397-88.474048	Waterbodies	Water body mask	Gneiss	Very low
16	Clayey	0-8.029984	59.665397-88.474048	Waterbodies	Water body mask	Charnockite group	Very low
17	Clayey	0-8.029984	59.665397-88.474048	Waterbodies	Water body mask	Charnockite group	Very low
18	Clayey	11.7-19.798347	150.827775-171.75126	Residential	Ridge type structural hills (large)	Charnockite Group	Very high
19	Clayey	11.7-19.798347	150.827775-171.75126	Residential	Ridge type structural hills (large)	Charnockite group	Very high
20	Clayey	11.7-19.798347	171.75126-200.559911	Agriculture	Ridge type structural hills (large)	Charnockite group	Very high
21	Clayey	11.7-19.798347	171.75126-200.559911	Land without scrub	Ridge type structural hills (large)	Charnockite group	Very high

$$P(y \mid x_1, \ldots, x_n) p(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \underset{y}{\operatorname{argmax}}\ P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

Proposed architecture: The new approach is to classify and predict the possibilities of landslides by using a novel Soft Bayesian Prediction Model (SBPM). The innovative data mining classification approach is applied to predict the occurrence of the landslide using the detected landslide locations.

To predict the possibilities of landslides in hills areas by using various systematic models are used. For these predict analysis models are used, so, many environmental geological attributes. In that some of the attributes are fully involved to predict class labels but few attributes are partially involved. To overcome this disadvantage, we developed a novel Soft Bayesian Prediction Model (SBPM) to predict the incidence of the landslide using the detected landslide locations and the constructed spatial database. The predicted landslide analysis results are verified using the landslide location test data for each studied area. The results obtain are verified for accuracy and further fine tuned using hybrid soft computing techniques (Fig. 1 and Table 1).

Proposed SBPM Algorithm 1:

Input Att: set of attributes, n: number of attributes
Output SBPM classifier:
BEGIN
1. Calculate approximation values for all attributes using equation:

$$\mu_{P,Z}(\alpha) = \frac{\text{card}(\text{POSp}(Z)) - \text{card}(\text{POSp}_{-(\alpha)}(Z))}{\text{card}(U)}$$

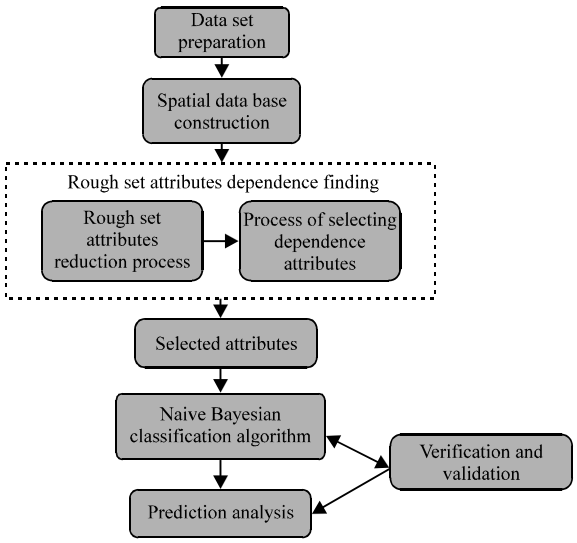

Fig. 1: Proposed frame work for Soft Bayesian Prediction Model (SBPM)

2. Compare the approximation values of each attribute aj
3. If both attributes are highly associated, i.e., ( ) remove the less approximation value Attribute(s).

4. For each j= i+1 To n DO

4.1. calculate $P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$ value for the all class labels

4.2. Compare the class labels value and predict the landslide occurrence

5. NEXT j
END

RESULTS AND DISCUSSION

Study area: In this research, we considered the study area is Coonoor of Tamil Nadu is painstaking the

analysis. Because landslides are occurring frequently in this area and landslide analysis has always been a concern here. As per fast historical data the landslide triggered due to the heavy rain occurred throughout the Coonoor and Ooty region of Tamil Nadu. "The landslide demolished nearly 300 tinned roof mud huts. Ketti and its border, about 7 km away from Ooty, received record rainfall of 820mm in 24 h while Ooty recorded 170 mm. As per another media report as many as 543 landslips occurred in just 2 days (10-11) in the Nilgiris and 816 houses razed to trash. Besides, 600 ha of crops have been devastated and road revetments damaged in 145 places. Above all, 43 precious lives lost and over 1,100 people have been left homeless". In addition, road accidents and traffic are a common problem on Coonoor and Ooty National Highway due to the regular landslips. In this study, we are tried to look into the given region of land in order to predict its susceptibility to landslip using soft Bayesian Prediction Model to classify the landslides (Fig. 2).

Experiment and result discussion: Soft Bayesian Model is produce the highest accuracy rate when compare to other data mining classification algorithms. This classifier can accept any number of either continuous or categorical variables. In fact the soft Bayesian Prediction Model classifier technique is particularly suitable for high dimension data. The soft Bayesian Prediction Model learning just reduces the probability of an inconsistent hypothesis. This gives the soft Bayesian Prediction Model learning a bigger flexibility.

There are six landslide factors considered in this analysis such as soil, slope, rainfall, land use and land cover, geomorphology and geology. The training dataset contains 445 record samples; each has 6 conditional attributes and 1 decision attribute. We have developed java codes for finding the rough set attribute importance approximation value for the different conditional attributes using soft computing techniques. All environmental attributes can be classified based on the attribute importance approximation value which we obtained through the computational program based on Table 2, the accuracy of the soil and geomorphology are very less when compared to the other conditional attributes. Based on the rough set core and reduct concept we remove both the attributes from the input knowledge data.

Table 2: Rough set attribute significance values

Name of the attribute	Attribute dependency approximation value
Soil	0.51
Slope	0.99
Rainfall	0.87
Land use and land cover	0.76
Geomorphology	0.56
Geology	0.68

The reduct data will be supplied to the Naive Bayesian algorithm to classify the landslides. The accuracy of soft Bayesian Prediction Model is 71.23% but when we applied the original data set to the naive bayesian without attribute reduction the accuracy of the result is 66.52%. The results have been shown in the graphs as follows Fig. 3-6.

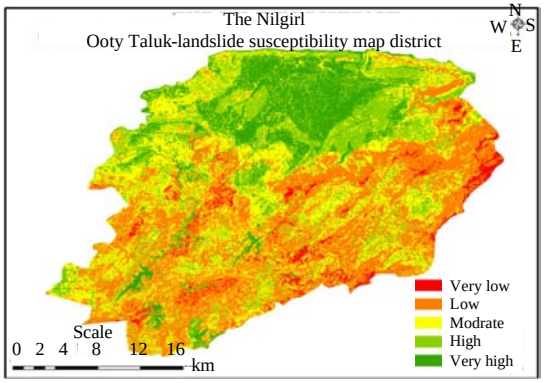

Fig. 2: Study area Ooty Taluk of Tamil Nadu

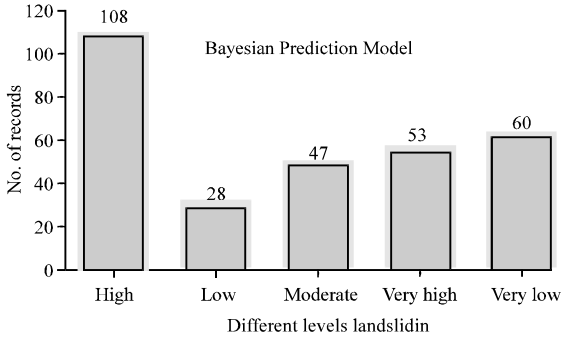

Fig. 3: Naive Bayesian without rough set attribute reduction

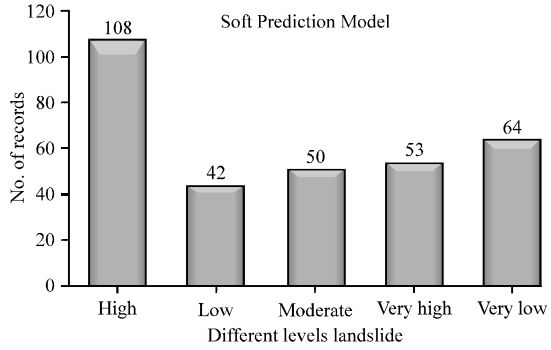

Fig. 4: Naive Bayesian with rough set attribute reduction (concidered highly significance attribute only)

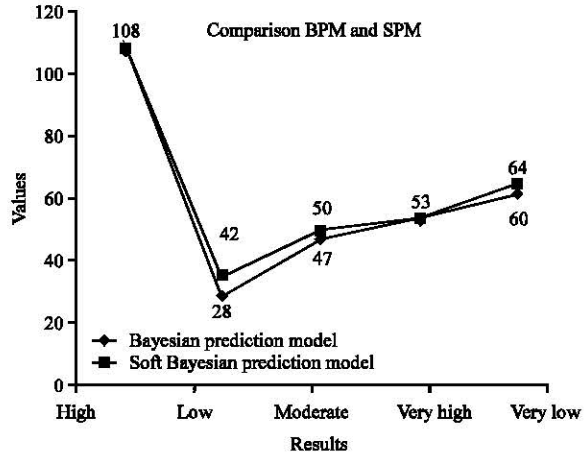

Fig. 5: Predicted values comparison

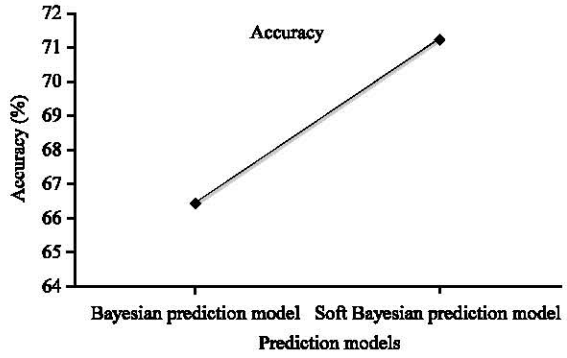

Fig. 6: Analysis of accuracy for Naive Bayesian and soft Bayesian Prediction Model

CONCLUSION

In this research, rough set and Bayesian classification approach is used to build novel Soft Bayesian Prediction Model (SBPM) to classify the landslides with the help of spatial database and GIS. We also compared the performance of soft Bayesian Prediction Model approach vs. Naive Bayesian classification. Soft Bayesian Prediction Model approach is more suitable and accurate than naïve bayesian. The method used in this study has both advantages and disadvantage for analyzing the link between landslide happening and environmental factors when compared with other methods. One advantage is the ability to analyze each factor or combination of factors probably affecting landslide happening independently without assuming an a priori model or relationship between environmental factors and landslide incidence. Another advantage is that these statistical tools can deal with ambiguity and vagueness in data and data mining is useful information from a large amount of data to classify.

RECOMMENDATIONS

In future research, more environmental factors will be considered for landslide prediction and also covering rough set techniques can be incorporated to avoid the uncertainty in the landslides. These results can be used as basic data to lend a hand to the slope management and land use development.

REFERENCES

Anbalagan, P. and R.M. Chandrasekaran, 2015. A novel weighted decision tree prediction model for landslide risk analysis. Adv. Nat. Appl. Sci., 9: 22-29.

Arciszewski, T. and W. Ziarko, 1990. Inductive learning in civil engineering: Rough sets approach. Comput. Aided Civil Infrastruct. Eng., 5: 19-28.

Beynon, M., 2001. Reducts within the variable precision rough sets model: A further investigation. Eur. J. Oper. Res., 134: 592-605.

Chung, C.J., 2006. Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment. Comput. Geosci., 32: 1052-1068.

Gorsevski, P.V. and P. Jankowski, 2008. Discerning landslide susceptibility using rough sets. Comput. Environ. Urban Syst., 32: 53-65.

Melchiorre, C., M. Matteucci, A. Azzoni and A. Zanchi, 2008. Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology, 94: 379-400.

Nefesliogl u, H.A., C. Gokceoglu and H. Sonmez, 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng. Geol., 97: 171-191.

Pawlak, Z. and R. Sowinski, 1994. Rough set approach to multi-attribute decision analysis. Eur. J. Oper. Res., 72: 443-459.

Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inform. Sci., 11: 341-356.

Pawlak, Z., 1997. Rough set approach to knowledge-based decision support. Eur. J. Oper. Res., 99: 48-57.

Rouse Jr., J.W., R.H. Haas, D.W. Deering and J.A. Schell, 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Progress Report RSC 1978-2, Texas A&M University, Remote Sensing Center, College Station, Texas.

Saito, H., D. Nakayama and H. Matsuyama, 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. Geomorphology, 109: 108-121.

Saxena, A., L.K. Gavel and M.M. Shrivas, 2014. Rough sets for feature selection and classification: An overview with applications. Intl. J. Recent Technol. Eng., 3: 62-69.

Slowinski, R., S. Greco and B. Matarazzo, 2012. Rough Sets in Decision Making. In: Encyclopedia of Complexity and Systems Science, Meyers, R.A. (Ed.). Springer, New York, USA., ISBN:978-1-4614-1799-6, pp: 2727-2760.

Wan, S., 2009. A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. Eng. Geol., 108: 237-251.

Wan, S., T. Lei and T. Chou, 2010. A novel data mining technique of analysis and classification for landslide problems. Nat. Hazards, 52: 211-230.

Wu, C.H. and S.C. Chen, 2009. Determining landslide susceptibility in central from rainfall and six site the analytical hierarchy process method. Geomorphology, 112: 190-204.

Yao, Y. and B. Zhou, 2016. Two Bayesian approaches to rough sets. Eur. J. Oper. Res., 251: 904-917.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio,, artificial neural networks and their comparison: A case study from Katlandslides (Tokat-Turkey). Comput. Geosci., 35: 1125-1138.

Zeng, Z.P., H.B. Wang, Z. Zhang and C.S. Xue, 2006. susceptibility assessment in the qingganhe river of three gorges area. Chin. J. Rock Mech. Eng., 25: 2777-2784.

Zhang, J., J.J. Jiao and J. Yang, 2000. In situ rainfall infiltration studies at a hillside in Hubei Province, China. Eng. Geol., 57: 31-38.