APPLICATION OF DATA MINING TECHNIQUE TO PREDICT LANDSLIDES IN SRI LANKA

Karunanayake K.B.A.A.M and Wijayanayake W.M.J.I.

Department of Industrial Management, Faculty of Science, University of Kelaniya,
Sri Lanka

ABSTRACT

Landslides are the major natural disaster in hill country of Sri Lanka, which create terrible economical and ecological damages. Therefore, the fast detection is important. Currently in Sri Lanka,predict landslides based on a map reading approach. But a map is limited to specific point in time, and do not take current conditions into account. Therefore, develop a model/tool which has ability to efficiently deal with current situation is important.

Hence within this study, prediction models were developed using Decision Tree and Neural Network data mining techniques,based on the data of Badulla and NuwaraEliya districts.

Selected Decision Tree model for Badulla district has 96.2963% accuracy level and Nuwara Eliya district has 100% accuracy level. Though Decision tree models were outperformed, Neural Network models also have above 90% accuracy. Therefore, it can be concluded that both data mining techniques are suitableto developlandslide prediction models for Sri Lanka.

KEYWORDS

Landslide, Data mining, Predictive analysis, Plan-Do-Check-Act, Decision tree

1. INTRODUCTION

Landslides are a common natural phenomenon in many parts of the world, especially in hilly or mountainous terrains. A landslide event is defined as “the movement of a mass of rock, debris, or earth (soil) down a slope (under the influence of gravity)”. The word “landslide” also refers to the geomorphic feature that results from the event[1].

In Sri Lanka during the last few decades, landslides occurred with increasing frequency in the hill country. The hill country, underlain by highly folded, fractured and weathered metamorphic rocks, has a high probability for landslides during heavy rainfall[1].

The incidences of landslides in Sri Lanka has increased mainly withinCentral Highland districts include, Badulla and Nuwara Eliya.Disasters due to landslide hazards have brought significant economic and social impact mainly to these two districts causing severe damages to life and property, the environment, and socio-economic life of the society[2]. Therefore, the fast detection plays an important role in avoiding or minimizing the hazards.In order to that different types of early warning mechanisms use in all over the world.

In Sri Lanka, the responsible body for landslide hazards, National Building Research Organization (NBRO) uses a map reading approach to determine the probability of landslides. NBRO issues landslide warnings based on Landslide Hazard Zonation Map and auto meter rain gauging established in important areas. But a map is only covering a specific point in time, and do

DOI:10.5121/ijdkp.2019.9404

not take current weather and geographical conditions into account. Though they collect current rainfall using auto meter rain gauging, this facility is not established in everywhere. As the hill country is a rapidly developing area the factors like land use and slope in a particular area can be change time to time. On the other hand, to deal with the current approach there must have an expert and it required more time to get a result.

Therefore, it is important to develop a model/tool which is very user friendly, efficient and can be deal with data of current situation. It is much useful having a tool even can be used by urban planners and developers to know the landslide potential of the target area, to plan the future developments in this hilly region economically and effectively. As well as it is important for each and every person who is living in a landslide prone area have an idea about, “am I safe in the current place with regards to current geological and weather condition?” Rather than living blindly until NBRO issue warnings.

During this study, mainly focused on develop a suitable predicting model using predictive analysis data mining techniques by combining most prominent and rapidly varying causative factors, slope range, land use and land overburden and triggering factor rainfall data and develop a user friendly and efficient software tool based on the developed model for the geographical context of Badulla and Nuwara Eliya districts.

2. RELATED WORK

In Sri Lanka during the last few decades, landslides occurred with increasing frequency in the hill country. The hill country, underlain by highlyfolded, fractured and weathered metamorphic rocks, has a high probability for landslides[1].

The intensity of a landslide is evaluated on the quality of debris load, number of deaths, number of affected families, value of damaged properties etc. Compared with floods, droughts, coastal erosion and cyclones, landslide appear as small incidences but badly affects life and property[3]. Therefore, the fast detection plays an important role in avoiding or minimizing the hazards.

The National Building Research Organization (NBRO) has proposed six causative factors after studying around 1700 landslides occurred in all around Sri Lanka[4]. The contribution or weightages of each of the six factors as a percentage are given in following table.

Table 1:Causative factors with weightages

Causative factors	**Weightage**
Bedrock geology and geological structures	20
Surface deposits/Overburden	10
Slope range	25
Hydrology and drainage	20
Land use and management	15
Landform	10

The causative factors such as bedrock geology, slope angle, landform, overburden soil cover considers as static factors and drainage pattern and land use patterns consider as dynamic factors mainly due to human intervention[5].

The rocks in the hill country are highly weathered and metamorphosed. When this weathered material is saturated with rain water, the mass of earth and rock to move rapidly down the slope under the force of gravity[2].
According to the above mentioned studies bedrock geology is static feature in Hill country area which formed in a way such that prone to occur landslides easily.

The hydrology and drainage pattern factor is vary due to human intervention[5]. Though it mentioned like that in some studies, when it comes to hill country area drainage pattern is static with the plantation of hilly areas[2]. As well as it is a poor drainage condition leading to excessive water seepage in sub strata[6].

The landform is a static factor for a specific geographical area[5]. As well as Central Hill Country characterized by Rolling and hilly terrain[2].

According to above studies there has evidence to conclude that as this study is relevant to a specific geographical area, a part of hill country, the causative factors bedrock geology, hydrology and drainage pattern and landform are relatively similar and static in long run.

The slope range has very high relative importance among the six causative factors[7]&[4] and it is a static factor [5]. But due to construction and cultivation practices slope range can be vary in smaller areas[8]. The slope surfaces with thick soil layers and slope angle between 15° and 45° have been found to have a greater preponderance for land sliding with maximum tendency of hill slope of angle 26° to 35° to the horizontal[6].

Slope surface also known as overburden also a causative factor for landslides in Sri Lanka. The main rock types igneous and metamorphic and these are deeply weathered in the hill country to give considerable thickness of residual soil as well as carpets of colluvium on foot slope[9]. The slope surface also a static factor[5]. But according to the definitions of residual and colluvium soils, it can be change time to time due to natural incidents as well as human intervention in smaller targeted areas[10]. The land use is the frequently dynamic factor due to constructions, agricultural activities or other environmental researches[5]&[8].

According to above studies there has enough evidence to conclude that, as the study is targeted on predicting the landslide riskiness of small targeted areas of part of the hill country area there has a probability of changing the slope range and surface overburden of target areas due to natural incidents or human intervention. Also, above studies provide evidence that due to the increase of human settlement in hill country area land use pattern is changed rapidly.Earthquake and rainfall are the main crucial triggering factors for landslide[11].

The landslides in Sri Lanka are generally followed intense and continuous rainfall exceeding a threshold between 350 to 400 mm within two days. The landslide hazard increases with the increasing rainfall received. Increasing intensity is known to trigger landslides and very high correlation is seen between the locations of the past landslides and the areas of increasing rainfall intensities[12]. Therefore, there has enough evidence to conclude that rainfall is the main triggering factor for the landslides occurring in the Sri Lanka.

Early-warning systems (EWSs) and other mechanisms used to detect and predict landslides are crucial to reduce the risk of landslide, especially where the structural measures are not fully capable of preventing the devastating impact of such an event.

National Building Research Organization (NBRO) is the responsible technical agency issue of early warnings for landslide hazards in Sri Lanka. In order to this National Building Research

Organization (NBRO) under the Ministry of Disaster Management in Sri Lanka has been conducting a Landslide Hazard Zonation Mapping in the ten districts of hill country areas since1989 covering entire districts of Badulla, Nuwara Eliya, Matale, Kandy, Kalutara, Ratnapura, Kegalle, Matara, Hambantota& Galle districts in 1:50000 & 1:10000 scale. Also they have established Auto meter rain gauging for important locations. This gives more accurate rainfall data for early warning. This can update every hour[13]. But these Landslide Hazard Zonation Maps only cover a specific point in time, and do not take current weather conditions into account[14].

In order to overcome from the above mentioned problem NBRO has introduced an auto meter rain gauging for important locations. But this is not for everywhere. On the other hand they do not concern about getting real time value of any geological factor mentioned above. Though most of the factors such as bedrock geology and geological structures, surface deposits, hydrology and drainage, landform remain same as this study concern limited area in hill country, but other two factors, slope range and land use can be change time to time due to human intervention as discussed above when considering a specific smaller area. And another thing is, to deal with this map and auto meter rain gauging there must have an expert. Therefor there is a gap of accurately predicting landslides with the real time data as well as this mechanism is not much user friendly. So it is very important to have a model which can be deal with current geographical data as well as which can be use by ordinary people by incorporating most contributing and varying causative factors, slope range and land use, as well as surface overburden which can be consider as constant factor, but in some points due to human intervention and natural incidents this can be vary when considering specific smaller area and major triggering factor rainfall.

Landslides often occur at specific location under certain topographic and geologic conditions within the country and it is important to utilize existing data to predict landsides[15]. In recent years, data mining approaches offer a fresh approach to analyze landslides and geosciences[11]. The definition for “Data Mining” is, analogously, data mining should have been more appropriately named “knowledge mining from data”[16]. According to the nature of landslide occurrence data and definition of the data mining it provides evidence data mining is a correct approach to develop a model to predict landslide riskiness of a particular area. The classifications and regression for predictive analysis is a major use of data mining techniques. Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts[16].

According to a study done in China, the decision tree analysis is used as for comparison with Discrete Rough Set (DRS) method. According to the results of their study DRS method of evaluating landslide occurrence which has an accuracy of 73%. This is somewhat similar to the result of conventional Decision Tree Method[11].

Artificial Neural Networks are generic nonlinear function approximates, extensively used for pattern recognition and classification, which also have a wide applicability in system identification. The ANN approach has many advantages compared with other statistical methods; it is an effective way for forecasting in complex nonlinear dynamic systems. In recent years, the ANN is widely used in the area of landslide forecasting and predicting[17].

A study done in Italy was aimed to compare Binary Logistic Regression (BLR) and Stochastic Gradient Treeboost (SGT) which is a decision tree method in assessing landslide susceptibility within the Mediterranean region for multiple-occurrence regional landslide events. According to the result of their study the SGT proved to be highly performing in small catchments; in general, this could suggest the application of this method for small catchments or areas with a dense distribution of homogeneous predictors. On the other hand, the BLR proved in bigger catchments to be more adaptive or less sensitive to predictor changes. This condition could lead to application of BLR in contexts of bigger catchments, where changes in the predictor spatial distributions will have a stronger control on the final model[18].

One of a related landslide prediction study has done using Naïve Bayes Theorem.But it was not much successful[19]. As well as Naïve Bayes Theorem assume that all attributes have same weight. It is a contradiction with concepts which are discussed at the factors affecting to landslides, because NBRO has provide different weightages for different factors. Therefore, Naïve Bayes Theorem is not suitable for this study.

According to the above findings the occurrence of landslides could be non-linear. If the occurrence of landslide is non-linear the models developed using neural network will be performed very well. Therefore, neural network is a suitable data mining technique for this study.
As well as based on the result of one of the above mentioned study, able to decide that, stochastic gradient treeboost (SGT) which is a decision tree method has ability provide highly performing models for small catchments; in general, this could suggest the application of this method for small catchments or areas with a dense distribution of homogeneous predictors. The main objective of this study to predict landslide riskiness of small catchment based on the current geographical and weather condition. As well as study scope is limited to Badulla and Nuwara-Eliya districts which have geographical homogeneity. Therefore, decision tree algorithm also suitable for this study.

Based on the above justification decision tree algorithm and neural network algorithm were used to develop models in this study out of classification rules for predictive analysis in data mining.

3. METHODOLOGY

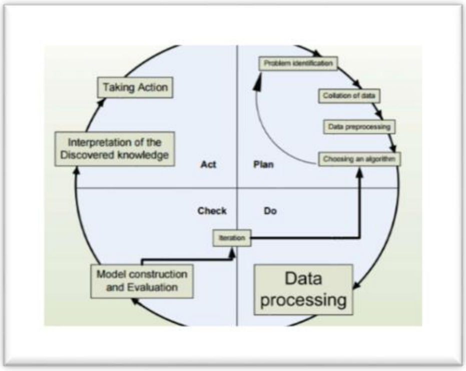


Figure 1:Plan-Do-Check-Act life cycle

Plan-Do-Check-Act life cycle was used as the methodology for this study.

3.1 Collection of Data

Causative factors related data of previously occurred landslides were collected from historical data sets which are in form of digitalized maps. These maps are created by utilizing around 20 years of historical data. Slope, overburden, land use of previously occurred landslides were collected by intersecting contours, map of overburden and map of land use with map of landslide which are belonged to particular area. Get readings of maps, intersection of maps and transformation of maps were done using the Arc GIS tool.

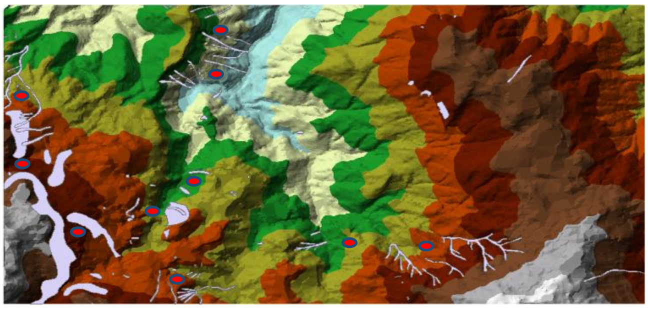


Figure 2: Map of landslide of Nuwaraeliya district intersect with Contour

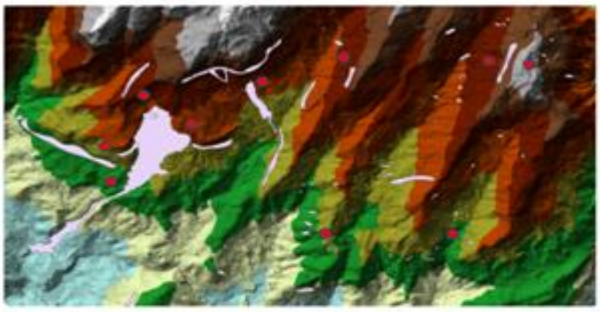


Figure 3:Map of landslide of Badulla district intersect with Contour

Triggering factor data (rainfall data) which are related to previously occurred landslides in Badulla and Nuwara-Eliya districts were collected based on the report, “Inventory of Landslides which have been occurred in the past in different districts” and findings of the research done in past which are relevant to the geographical context of the study. Based on the date of occurrence missing rainfall data were calculated using below graphs.

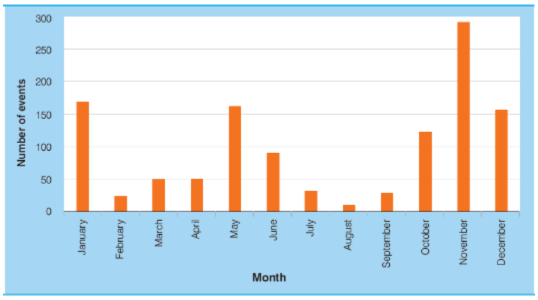


Figure 4:Graph of number of events respective to month

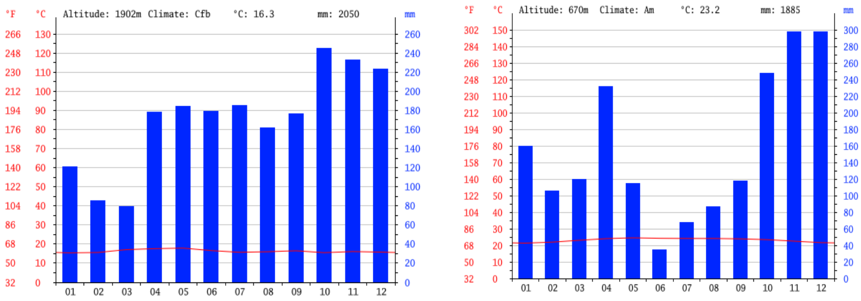


Figure 5:Average monthly rainfall for Nuwaraeliya and Badulla districts

Identified several points which are closed to occurred landslides as nonaffected points and causative factor details related to those points were collected using same method and rainfall data were calculated based on the rainfall value of nearest occurred landslide.

At the end of the process of data collection two data sets were built as follows.

Badulla District

81 records on occurred landslides + 10 records on not affected points= 91 data records

Nuwara Eliya District

81 records on occurred landslides + 10 records on not affected points= 91 data records

3.2 Data Pre-Processing

Collected data from various sources needs to be organised and cleaned such that a data mining technique can be readily applied. In this study there has two categorical predicting variables out of four, named, surface overburden and land use. The uniform categorization established based on the categorization used by NBRO.

Table 2: Categorization of land use types

Type	**Factor class**
Type 1	Coconut, Forest, Well managed tea
Type 2	Paddy, Rubber, Scrub, Home Garden, Poor managed tea
Type 3	Water bodies, Grasslands, Buildings

Table 3:Categorization of overburden types

Type	**Overburden**
NM	Bare bedrock
Coll	Colluvium soil
Rs	Residual soil
RE/Rs	Rock Exposure with Residual soil
RE/Coll	Rock Exposure with Colluvium soil

3.3 Algorithm Selection

Basedon thefindings of literature review Classification and regression for predictive analysis was identified as the suitable data mining technique for this study. Under this technique Decision tree and Artificial Neural Network algorithms were selected based on the matching of characteristics of each algorithm and results of the previously conducted similar researches. Two types of attributes were used in classification data mining techniques. Named as classifying attribute and predicting attributes. The classifying attribute must be a categorical attribute and predicting attributes can be numerical or categorical. In this study predicting attributes were slope in degree, overburden type, land use type and rain fall in mm. The classifying attribute was whether there has a landslide risk or not.

3.4 Data Processing

Data splitting was done in this stage. The data sets were divided in to two parts separately, such as 70% of each data set as training data and 30% of each data set as test data.

3.4.1 Data processing with ANN

In order to train neural network this data set should normalized. Normalization implies that all values from the data set should take values in the range from 0 to 1.

Following formula was used to normalized numerical data:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}$$

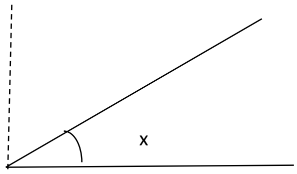
Where:

X –value that should be normalized
Xn – normalized value
Xmin – minimum value of X
Xmax – maximum value of X [20]

Following table shows the maximum and minimum rainfall values for each district[21]&[22].

	Xmin(mm)	**Xmax(mm)**
Badulla	35	300
Nuwara Eliya	79	250

The slope values taken from contours were horizontal angle of the slope with the earth.



Following table shows the maximum and minimum slope values for each district.

	Xmin(^{0})	**Xmax(^{0})**
Badulla	0	90
Nuwara Eliya	0	90

At the network design for each category there must be used separate node for each category. Therefore, during the data pre-processing separate column was used for each category. Data were represented in such a way that, if value is belonged to particular column represent it by 1 and if not represent it by 0.

The neural network models were developed using **"Neuroph Studio"** data mining tool. As this study was conducted as a supervised learning study feedforward neural network technique was used for neural network model implementation. During the feedforward neural network implementation data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs. During normal operation, that is when it acts as a classifier, there is no feedback between layers[23]. AndMulti-Layer perceptron (MLP) is a feedforward neural network algorithm in Neuroph studio with one or more layers between input and output layer[24].Therefore, neural network models were implemented using Multi-Layer Perceptron algorithm.

3.4.2 Data processing with Decision Tree

Since Decision Tree algorithms able to process data with both categorical and numerical data, data transformation was not needed for process data with Decision tree algorithm. The decision tree models were developed using **"Weka"** data mining tool. As this study was needed establishment of classification rules, visualization of decision tree models was a must. Therefore, J48 algorithm and random tree algorithm were used to develop decision tree models.

WEKA implements decision tree C4.5 algorithm using "J48 decision tree classifier". C4.5 has an enhanced method of tree pruning by replacing the internal node with a leaf node thereby reducing misclassification errors due to noise or too many details in the training data set[25].

Random tree models have been extensively developed in the field of machine learning in the recent years[25].

3.5 Model construction and evaluation

The predictive accuracy of the models were estimated using the test sets, made up of test tuples and their associated class labels. These tuples were randomly selected from the general data set. They are independent of the training tuples, meaning that they were not used to construct the classifier.

Receiver Operating Characteristic (ROC) is a standard method to evaluate general performances of models[18]. This method was used to evaluate the performance of each model. In statistics, a ROC, or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings[18].

Also, model implications will be compared with existing concepts related to landslide occurrence for higher accuracy.

4. Experimental Result

At the end of the data pre-processing there has been identified 2 separate data sets for data processing.

1. Badulla Data set
2. Nuwara Eliya Data set

Before process the data using data mining techniques each data set was split in to 2 groups such as, 70% as training set and 30% as testing set. This splitting was done in a random manner.

Table 4:No.of records of divided data sets

	Total	Training set	Testing set
Badulla	91	64	27
Nuwara Eliya	91	64	27

4.1 Artificial Neural Network

Following table shows a sample normalized data set used for process using ANN technique.

Table 5: Part of normalized data set used for ANN technique

Overburden					**Land Use**			**Slope**	**Rainfall**	**Risk/Not**
Rs	Coll	RE/Rs	RE/Coll	NM	Type1	Type2	Type3			
1	0	0	0	0	0	1	0	0.211111	0.924528	1
1	0	0	0	0	0	0	1	0.4	0.596226	1
0	1	0	0	0	0	0	1	0.288889	0.603774	1
0	1	0	0	0	0	1	0	0.211111	0.366038	0
1	0	0	0	0	0	0	1	0.211111	0.339623	0

As this study was conducted for classification data modelling the “feedforward” Neural Network algorithm was used to implement neural network models. In this study, “sigmoid” function used as the transfer function in each layer, because data sets were normalized and data were between 1 and 0. As well as bias neuron also used for neural network models, because the Bias weights control shapes, orientation and steepness of all types of Sigmoid functions through data mapping space. During the training process, the connection weights of the neural network were initialized with some random values. The “backpropagation” learning rule was used to adjust the connection weightages according to error.

Data processing with Neural Network data mining technique was done using the Neuroph Studio software tool with a nature of 0.01 Maximum error, 0.2 learning rate 0.2 and 0.7 momentum.
Trial and error approach was used to find suitable number of hidden layers and nodes. In order to that the experimental approach was conducted on several neural network architectures to identify the best architecture for each data set by changing the number of hidden layers and number of hidden neurons. Following are some of the criteria which were used for selecting the number of hidden neurons.

1. The number of hidden neurons should be in the range between the size of the input layer and the size of the output layer.

2. The number of hidden neurons should be the input layer size, plus the size of the output layer plus one.
3. The number of hidden neurons should be less than twice the input layer size

Other than these 3 rules experiment was also conducted for random number of hidden neurons. As well as number of hidden layers were limited to 1,2 and 3 as commercial products only have capability to process data with maximum 3 layers.

Based on the above-mentioned rules and conditions highly performed neural network architecture for Badulla district data set was consist with 3 hidden layers and 1 hidden neuron per each layer. Similarly, highly performed neural network architecture for Nuwara Eliya district data set was consist with 1 hidden layer and 10 hidden neurons per each layer.

Following table describes a summary of training and testing results of selected architectures.

Table 6:Smmerized result of ANN technique models

	Total Square Mean Error (TMSE) of training	Total Square Mean Error (TMSE) of testing	Correctly Classified Instances (CCI) - Level of Predictive Accuracy	
			Proportion	Percentage
Badulla data set	0.0156	0.0741	25/27	92.5926%
Nuwara Eliya data set	7.99E-08	0.037	26/27	96.2963%

According to the above results though the neural network model which was developed using Nuwara Eliya District data set has a good CCI percentage its TMSE for training value is very less. This might be led to overfitting of the model. The over fitted models' predicting capability is less. Therefore, the neural network model which was developed using Nuwara Eliya district was not selected for further analysis. The neural network model which was developed using Badulla district was selected for further analysis.

4.2 Decision Tree technique

Following table shows a sample pre-processed data set used for process using Decision Tree algorithms.

Table 7:Sample of pre-processed used for Decision Tree technique

Overburden Type	**Land Use Type**	**Slope (^{0})**	**Rainfall (mm)**	**Risk/Not**
Rs	Type2	19	280	RISK
Rs	Type3	24	218	RISK
Coll	Type3	26	195	RISK
RE/Coll	Type3	25	196	RISK
RE/Rs	Type3	13	209	RISK
NM	Type1	33	217	NO RISK
Rs	Type3	19	125	NO RISK

Categorical Numerical Classifying attribute

The J48 and random tree algorithms in WEKA tool were used to develop the Decision Tree models. Following table describes the capabilities of Decision Tree models.

Table 8:Summary of Decision Tree models

	J48			Random tree		
	Mean Absolute Error (MAE) for testing	Correctly Classified Instances (CCI) - Level of Predictive Accuracy		Mean Absolute Error (MAE) for testing	Correctly Classified Instances (CCI) - Level of Predictive Accuracy	
		Proportion	Percentage		Proportion	percentage
Badulla data set	0.0695	26/27	96.2963%	0.0412	26/27	96.2963%
Nuwara Eliya data set	0.0536	26/27	96.2963%	0	27/27	100%

Based on the CCI percentages and MAE values Random tree models were outperformed than J48 models for both data sets. Therefore, Random tree models were selected for further comparisons. Following table shows the summary of the selected models of two data mining techniques according to the results of experiments.

Table 9:Summary of selected models

	Correctly Classified Instances (CCI) percentage for testing – Level of Predictive Accuracy	
	Decision tree model	Neural network model
Badulla district	96.2963%	92.5926%
Nuwara Eliya district	100%	Over fitted model

According to the above summarized CCI percentages, decision tree models were out performed than the neural network models. Therefore, random tree Decision Tree models have the highest level of predictive accuracy. Hence these Decision Tree models were selected as the best suitable models for predicting landslide riskiness of a smaller targeted area of both Badulla and Nuwara Eliya districts according to the current weather and geographical conditions.

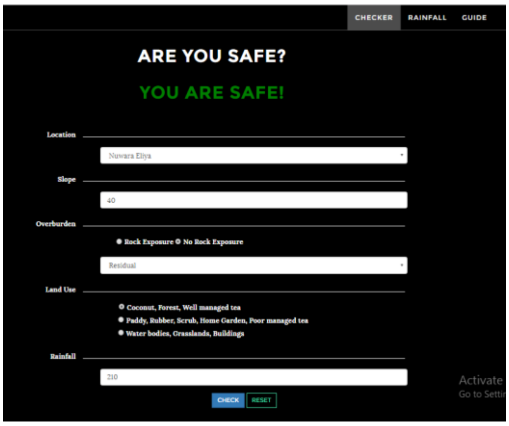
According to the structures of the selected models “Rainfall” factor has the highest priority among four predicting attributes. This factor is correctly mapped with the theory behind landside occurrence in Sri Lanka. Also, other causative factors are in the correct order according to the theories behind the causative factors affect to the landslides in Sri Lanka.

Based on the above mentioned CCI percentages of model testing, these models’ predictive accuracy has higher level of validity. Moreover, according to [Lombardo, et al., 2014] Receiver Operating Characteristic (ROC) is a standard method to evaluate general performances of models. The area under ROC curve is determined by the true positive and false positive rate. ROC area = 1 indicates a perfect prediction, while ROC area = 0.5 indicates a random prediction. According to the observation both true positive and false positive rates of NuwaraEliya district decision tree model equals to 1 while ROC area equals to 0.5 for both classes of Badulla district decision tree model. According to the definition of ROC values NuwaraEliya district model has perfect prediction and Badulla district model has random prediction.

Though Decision Tree model for Badulla district has been implied a random prediction nature at the class wise evaluation as a whole predictive accuracy level of this model at a higher level.
The decision tree model which was developed using Nuwara Eliya district data set's ROC area is equalled to 1 and it has a 100% predictive accuracy level.

Therefore, both decision tree models have higher level of accurately predicting capability as well as both models were followed the correct order factor weightages.

As an outcome of this study at the end, a simple prototype web application named "Landslide Early Warning System" was developed using derived classification rules of selected decision tree models. These classificationrules are based on static data sets. By providing the required data to the web form, the application will display the predicted result, whether the selected small catchment has risk of landslide occurrence or not,according to the given values for causative factors and triggering factor. This application could be enhanced as an advanced decision support systemby providing dynamic data sets to train the models.



ARE YOU SAFE?

YOU ARE SAFE!

Location

Nuwara Eliya

Slope

40

Overburden

● Rock Exposure ◉ No Rock Exposure

Residual

Land Use

◉ Coconut, Forest, Well managed tea
● Paddy, Rubber, Scrub, Home Garden, Poor managed tea
● Water bodies, Grasslands, Buildings

Rainfall

210

Figure 6:An interface of web application prototype

5.CONCLUSION

This study was conducted to develop a model to predict landslide riskiness of small catchments in hill country of Sri Lanka by considering current weather and most significant and dynamic geographical conditions of the particular area. In order to accomplish this objective, the prediction models were developed using Decision Tree data mining technique and Artificial Neural Network data mining technique which are belong to data mining for predictive analysis category. Models were developed by incorporating triggering factor – Rainfall and three causative factors named, slop, land use and form out of six causative factors.

For both data sets – Nuwara Eliya district dataset and Badulla district dataset, Random tree Decision tree models were outperformed than the other Decision Tree models and Neural network models. Selected Badulla district models has 96.2963% of predictive accuracy level while Nuwara Eliya model has 100% of predictive accuracy level.

But according to the results of this study there is only slight increase of predictive accuracy level percentage in Decision Tree models than the Neural Network models. The outperforming of these two data mining techniques might vary based on the district as well as set of selected factors. Therefore, it can be concluded that both decision tree technique and neural network technique are suitable to develop landslide riskiness predicting models. Furthermore, the same methodology and approach can be applied to any district in Sri Lanka to develop landslide prediction models.

REFERENCES

[1] A. M. K. B. Abeysinghe, Y. Iwoa, A. Saito and R. M. S. Bandara, "Hazard and Risk Assessment in Landslide Prone Hill Country of Sri Lanka," Geotechnical Journal, 2005.

[2] J. Katupoth, "Landslides in the Sri Lanka in 21st Centry," 1994.

[3] J. Katupotha, "EFFECT OF LANDSLIDE ON SOCIETY AND," Singapore, 2015.

[4] A. Silva and S. Kumara, "A case study in Diyadawa, Deniyaya," Landside Hazard Zonation Using GIS Techniques, 2005.

[5] M. Somarathne, "An Overview of Koslanda Landslide," 2014.

[6] R. M. S. Bandara, "Landslides in Sri Lanka," Vidurava, vol. 22, no. 2.

[7] K. M. Weerasinghe, M. Gunerathne, H. G. P. A. Rathnaweera, U. G. A. Puswewala and N. M. S. I. Armbepola, "Analytical Determination of Landside Potential Using Fuzzy Sets and Other Statistical Technics".

[8] L. Zubair, V. Ralapanawa, U. Thennakoon, Z. Yahiya and R. Perera, "Mapping Hazards and Risk Hotspots," Natural Disaster Risk in Sri Lanka, 2003.

[9] S. Singh and S. Singh, "Occurence and Significance of Landslides in Southeast Asia," in Disaster management, 1998.

[10] N. B. R. O. NBRO, Hazard Resiient Housing Construction Manual, 1st ed., National Building ResearchOrganization, 2015.

[11] S. Wan, T. C. Lei and T. Y. Chou, "A noval data mining technique of analysis and classification for landslide problems," Springer Science + Business Media B.V, 2009.

[12] U. Rathnayake and S. Herath, "Changing rainfall and its impact on landslides in Sri Lanka," Journal of Mountain Science, 2005.

[13] P. A. N. UN-Asia, "Early Warning Systems for Disasters in Sri Lanka," 2014.

[14] I. o. E. S. E. Fraunhofer, "An accurate way of predicting landslides, Research News," 2013. [Online]. Available: https://www.fraunhofer.de/en/press/research-news/2013/march/an-accurate-way-of-predicting-landslides.html.

[15] "Landslides in Japan, 5th revision.," [Online]. Available: http://web.tuat.ac.jp/~sabo/lj/ljap3.htm. [Accessed 20 9 2016].

[16] J. Han, M. Kamber and J. Pei, Data mining concepts and techniques, 2012.

[17] H. Chen and Z. Zeng, "Deformation Prediction of Landslide Based on Improved Back-Propagation Neural Network," Springer Science + Business Media LLC, 2012.

[18] L. Lombardo, M. Cama, C. Conoscenti, M. Marker and E. Rotigliano, "Binary logistic regression versus stochastic gradient boosted decision trees in assessing landslide susceptibility for multiple-occurring landslide events: application to the 2009 storm event in Messina (Sicily, southern Italy)," Springer Science+Business Media Dordrecht, 2014.

[19] B. T. Pham, D. T. Bui and I. Prakash, "Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS," Journal of Geomatics, vol. 10, no. 1, 2016.

[20] neuroph.sourceforge.net.. [Online]. Available: http://neuroph.sourceforge.net/tutorials/wines1/WineClassificationUsingNeuralNetworks.html. [Accessed 26 12 2016].

[21] "en.climate-data.org.," [Online]. Available: https://en.climate-data.org/location/909/. [Accessed 26 12 2016].

[22] "en.climate-data.org.," [Online]. Available: https://en.climate-data.org/location/764256/. [Accessed 26 12 2016].

[23] "fon.hum.uva.nl.,"[Online].Available: http://www.fon.hum.uva.nl /praat/manual/Feedforward_neural_networks_1__What_is_a_feedforward_ne.html . [Accessed 26 12 2016].

[24] "http://neuroph.sourceforge.net," [Online]. Available: http://neuroph.sourceforge.net/tutorials/MultiLayerPerceptron.html . [Accessed 26 12 2016].

[25] S. S. V. R. C. Ravichandran, "Comparative Study on Decision Tree Techniques for," Journal of Communication and Computer, 2012.

AUTHORS

Ms. K.B.A.A.M. Karunanayake holds a Bachelor's(hons.) Degree in Management and Information Technology from Faculty of Science, University of Kelaniya, Sri Lanka and Professional Graduate Diploma from British Computer Society. Also, she is an Oracle Certified Associate, Java SE 7 Programmer. Her research interests are Data Engineering and Software Engineering. She is currently serving as a Software Engineer at GT Nexus Services (Pvt) Ltd. – Infor Sri Lanka.

Prof. W.M.J.I. Wijayanayake received a PhD in Management Information Systems from Tokyo Institute of Technology Japan in 2001. He holds a Bachelor's degree in Industrial Management from the University of Kelaniya, Sri Lanka and Master's degree in Industrial Engineering and Management from Tokyo Institute of Technology. He is currently a Professor at the Department of Industrial Management, University of Kelaniya, Sri Lanka. His research interests are Information System, Data Engineering, Enterprise Architecture, Software Engineering, and Business Intelligence. He also have worked as an IT consultant and provided advice to small and medium size enterprise managers on MIS development, Productivity Improvement etc.