Landslide susceptibility mapping using geospatial analysis and Recurrent Neural Network (RNN)

Quoc Phi Nguyen[1,*]
[1]*Hanoi University of Mining and Geology*

ABSTRACT

Landslide susceptibility map is an imperative basic tool for land use application, spatial planning and disaster mitigation. The landslide was governed by the interaction of several factors: lithology, slope angle, fractures inside rockmass, weathering profile, groundwater condition, rainfall, etc. This research examines the effectiveness of using a special architecture of Artificial Neural Network (ANN) in landslides predictions called Recurrent Neural Network (RNN). Geographic Information System (GIS) was applied to analyze the relation between landslide causative factors together with the landslide inventory and produce the causative factors maps. Then a RNN model was developed to characterize landslide-prone areas and the accuracies of the RNN model were compared with a Feed-forward Neural Network (FNN). The average success rates from RNN and FNN for the training samples were 87.53% and 79.46%, respectively. The accuracies of the RNN and FNN models for the test area were 83.68%, and 76.85%, respectively. The results show that the prediction accuracy of RNN model is much higher than the FNN for landslides along road corridors from Lao Cai province. The results suggest that, the RNN model is an effective and feasible method to further improve accuracy for landslide prediction in study area. The landslide susceptibility map can be used for preliminary landslide hazard prevention and mitigation, and proper planning for landuse and infrastructural development purpose.

Keywords: landslide susceptibility, Recurrent Neural Network, data mining, Lao Cai

1. Introduction

Landslides are dangerous geological disasters with catastrophic effects on human lives and properties. Vietnam is one of the areas where suffer the most serious geo-hazards such as landslides and debris flows in Asia. Landslide can be defined as a geological phenomenon under the influence of gravity, which can occur in offshore, coastal and onshore environments. Landslide includes a wide range of ground movements, such as rockfalls, deep failure of slopes and shallow slides. In the northern part of Vietnam, the large-scale landslides are notable for their scale and serious destruction. The last decades have witnessed an increase in the magnitude and frequency of these catastrophes. This is probably due to the rapid economic development and high population growth of Vietnam, which results in overexploitation of the environment including increased deforestation and occupation of mountainous and hillside areas.

In Lao Cai province, landslide is widespread and recurrent phenomena due to its particular geological and geomorphological patterns. Historical investigations have revealed that, in the last 10 years, at least 45 single and multiple large-scale landslides have caused 78 deaths and injured people. This trend will continue in future under increased unplanned urbanization and development, continued deforestation and increased regional precipitationin landslide prone areas. Therefore, the prediction of landslide is essential for carrying out quicker and safer mitigation programs, as well as future planning of the area. However, the prediction of landslide is a difficult task and requires a thorough study of past activities to determine the change condition. The stability of natural or manmade slopes was governed by the interaction of several factors, such as: lithology, weathering profile, geological engineering and hydrogeological conditions, drainage network, slope angle, landcover/landuse, etc., and hence, there has been a growing interest in questioning relationship between landslide hazard and related variables.

The Artificial Neural Networks (ANNs) is effective in geoscience applications as they are extensively used for pattern recognition and classification. From a structural point of view, ANNs can be classified into two main types: feedforward neural networks (FNNs) and recurrent neural networks (RNNs). In recent

**Corresponding author*
Email: nguyenquocphi@humg.edu.vn

years, the ANNs have been widely used in modeling the landslide susceptibility. However, the previous works mainly focus on the FNNs, the RNNs have not been applied in the area of landslide prediction so far. Previous studies show that ANN-based classifiers have been successfully in landslide prediction (Chang and Liu, 2004; Chen and Tang, 2015; Nguyen Quoc Phi et al., 2016), landslide detection and classification (Gorsevski et al., 2016: Chang et al., 2010), and landslide susceptibility (Dieu Tien Bui et al., 2016a, Dieu Tien Bui et al., 2016b, Quoc Phi Nguyen et al., 2018). However, the previous works mainly focus on the FNNs, the RNNs have not been applied in the area of landslide susceptibility mapping so far. RNNs are more appropriate for presenting nonlinear dynamic systems than FNNs (Lee H. and Park Y., 1991; Yu W., 2004). Since the landslides are essentially nonlinear dynamical systems, using RNNs for modeling can anticipate their behaviors more accurately. For this purpose, current research presents a framework to establish a landslide susceptibility model based on RNNs. The method was implemented for the prediction of landslides along road corridor of Lao Cai province, Vietnam.

2. Methodology

2.1. Artificial Neural Networks (ANNs)

The Artificial Neural Network is made up of a large number of independent inter-connected units, which are neurons and synapses. Upon receiving sufficiently intense stimulus (input) from the preceding units, the unit is activated and sends signal to the connecting units.

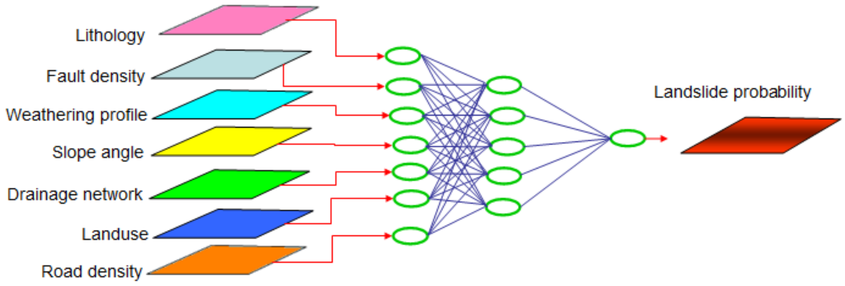

Figure 1. An example of landslide analysis in ANNs

The transformation is completed in two phases (Lee et al., 2007): Firstly, each input signal is multiplied by the weight of the connection and the results of the single products are added to obtain an amount called total input. Secondly, the unit applies a transfer function which transforms the sum of the input signals into output signals. The behavior of an ANN depends on the architecture of the network and on both the weights assigned to the connections and the transfer function.

2.2. Recurrent Neural Network (RNN)

A recurrent neural network (RNN) is a class of ANNs where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behavior for a time sequence. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. They are very powerful in computational analysis and are biologically more reliable than other NN techniques given their lack of internal states. The memory of past activations in RNN is very effective with feedback connections, making them suitable for learning the temporal dynamics of sequential data. RNN is very powerful when used to map input and output sequences because it uses contextual information. However, traditional RNNs face the challenge of exploding or vanishing gradients. Hochreiter and Schmidhuber (1997) proposed long short-term memory (LSTM) to tackle this issue. Hidden units in LSTM are replaced with memory blocks that contain three multiplicative units (input, output, forget gates) and self-connected memory cells to allow for reading, writing, and resetting through a memory block and behavioral control. A single LSTM unit is shown in figure below. c_t is the sum of inputs at time step t and its previous time step activations. LSTM updates time step i given inputs x_t, h_{t-1}, and c_{t-1} (Donahue et al., 2015).

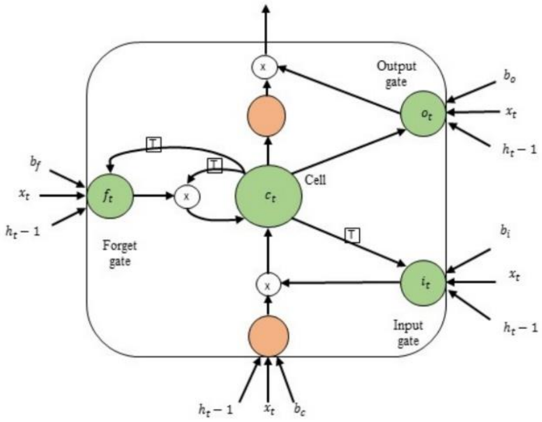

Figure 2. Structure of a memory cell in long short‐ term memory (LSTM)‐ RNN

Input gates:

$$i_t = \sigma(W_{xi}.x_t + W_{hi}.h_{t-1} + W_{ci}.c_{t-1} + b_i) \quad (1)$$

Forget gates:

$$f_t = \sigma(W_{xf}.x_t + W_{hf}.h_{t-1} + W_{xf}.x_t + W_{cf}.c_{t-1} + b_f) \quad (2)$$

Cell units:

$$c_t = i_t.\tanh(W_{xc}.x_t + W_{hc}.h_{t-1} + b_c + b_f.c_{t-1}) \quad (3)$$

Output gates:

$$o_t = \sigma(W_{xo}.x_t + W_{ho}.h_{t-1} + W_{co}.x_t + b_o) \quad (4)$$

The hidden activation (output of the cell) is also given by a product of the two terms:

$$c_t = o_t.\tanh(c_t) \quad (5)$$

where σ and *tanh* are an element-wise non-linearity, such as a sigmoid function and hyperbolic tangent function, respectively; *W* is the weight matrix; x_t refers to input at time step *t*; h_{t-1} represents the hidden state vector of the previous time step; and b_c denotes the input bias vector.

The memory cell unit c_t is a sum of two terms: the previous memory cell unit c_{t-1}, which is modulated by f_t and c_t, a function of the current input, and previous hidden state, modulated by the input gate i_t due to i_t and f_t being sigmoidal. Their values range within [0, 1], and i_t and f_t can be considered as knobs that the LSTM learns to selectively forget its previous memory or consider its current input, whilst o_t is an output gate that learns how much of the memory cell to transfer to the hidden layers.

3. Case study

3.1. Landslide database

A database for landslide susceptibility in study area was developed based on:

- The field surveys, were carried on with high accuracy GPS on topographic maps of 1:50,000 (partly 1:10,000). Historical data in study area were also collected from previous studies and reports.
- The geological maps of 1:200,000 (Bac Quang and Lao Cai-Kim Binh sheets), and recent 1:50,000 scale maps cover the study area were used.
- Topographic maps of 1:50,000, produced in 2004.
- Satellite images of Landsat 7 (ETM+) and Landsat 8 (OLI), which include scenes of 128/44 and 128/045 (path/row), taken in 2005, 2010 and 2015 with cloud cover <10% were used.

The landslide-related parameters were then prepared as shown in table 1.

Table 1. Derived variables utilized for landslide analysis

Components	Variables	Data sources	Map scales
Geological conditions	Lithology	Geological maps	1:200,000 1:50,000
	Geological structures	Geological maps, RS images	1:200,000 1:50,000
	Weathering profile	Geological maps	1:50,000

	Geological engineering conditions	Geological engineering maps	1:50,000
	Hydrogeological conditions	Hydrogeological maps	1:50,000
Natural conditions	Elevation	Topographic maps	1:50,000
	Slope angle	Topographic maps	1:50,000
	Slope exposure	Topographic maps	1:50,000
	Drainage networks	Topographic maps, RS images	1:50,000
	Land cover	Topographic maps, RS images	1:50,000
Human-induced conditions	Landuse	Topographic maps, RS images	1:50,000
	Road density	Topographic maps, RS images	1:50,000
	House density	Topographic maps, RS images	1:50,000

Identification of the landslide locations and delimitations was carried out by fieldwork, supported by analysis of the satellite images and historical data. The geo-database, collects information related to 82 landslide bodies along two main roads from the main city of Lao Cai (Laocai City) toward its adjacent county (Bao Thang).

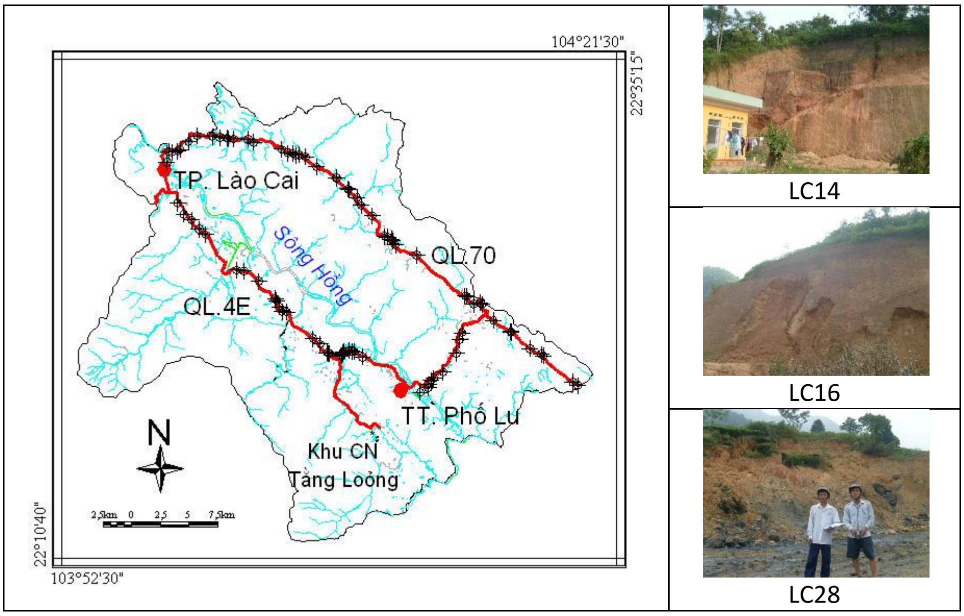

Figure 3. Landslide location map

Factors are grouped based on geological, natural and human-induced conditions (table 1). Hence, in addition to making a landslides inventory in the investigated area, the 13 variables described in table 1 were used as causal factors. The study was carried out using a 30m grid size DTM.

3.2. Analysis and Results

For the RNN, the work is divided in two steps: the training phase, in which two third (66%) of the total dataset were selected as training set and the weights were calculated, and the validation procedure, in which the obtained susceptibility map was cross-validated with the inventory database. The RNN model were implemented using the open source Orange3 deep learning framework.

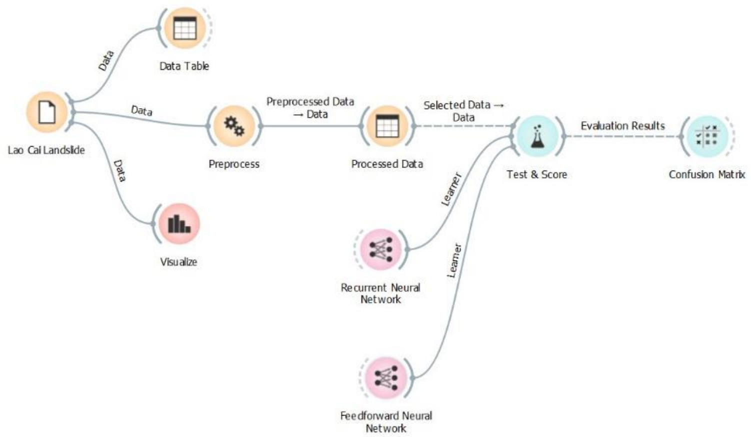

Figure 4. Workflow for landslide susceptibility analysis

In this study, RNN received 13 features as inputs to differentiate landslide and non-landslide locations. RNN consisted of an LSTM layer with 100 hidden units, two fully connected layers, a dropout layer, a softmax layer and an output layer showing the presence/absence of landslides. The back-propagation technique was used in trained the RNN model with Rectified Linear Unit (ReLU) activation and Stochastic Gradient Descent (SGD) solver.

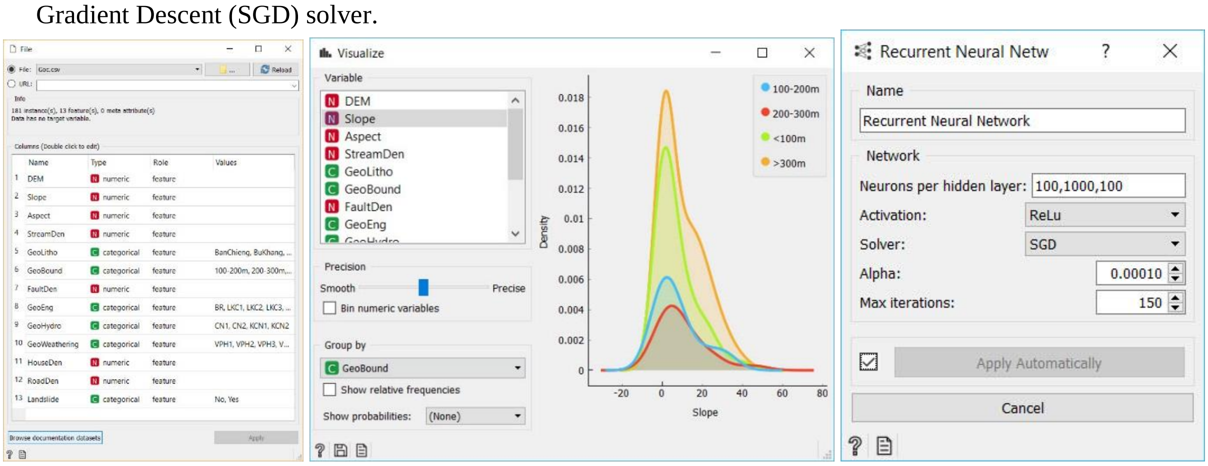

Figure 5. Input data visualize and optimized parameter for RNN model

After 150 iterations of the initial values, the optimal values for RNN model were obtained for landslide and non-slide areas and among the evaluated parameters, the model with the highest validation accuracy was selected. The qualitative assessment of the RNN model yielded high-quality results, landslide locations were detected and correctly differentiated from non-landslide locations with 87.53% for the training set and 83.68% for test set. On the other hand, the qualitative assessment of FNN produced low-quality results with 79.46% for the training set and 76.85% for test set. Overfitting can be avoided when dropouts are controlled through the number of parameters in the RNN model. The sensitivity analysis of the effects of dropout rate with various keep probability parameters on the RNN model. The results showed that the appropriate dropout rate is 0.6 for the RNN model.

4. Conclusion

The RNN modeling which is very good in learning and processing information can work out the complex nonlinear relation by using the present data to learn models. This paper has investigated the use RNN for landslide prediction. The RNN model exhibited better accuracy in the analysis and both training and test sets than the FNN model. Obtained results show that the RNN approach is effective at predicting landslide susceptibility along road networks in Lao Cai province. Therefore, this method has a good perspective in application and further development.

The susceptibility map prepared in this study is a step forward in the management of landslide hazard in study area. The quantification of landslide locations along road sections where landslides are small in size but occur frequently. Susceptibility assessment including small-sized landslides is important if these occur as even small slips can result in fatal accidents. Although the limited availability of information of landslide occurrence, especially the time span of collected data and exact date of landslides, the calculation of hazard maps was still feasible. To bring the prediction into real life, the combination of landslide hazard maps with additional information, such as travel distance, vulnerability assessment… is currently conducting for subsequent risk analysis and estimates the probable losses due to future landslide events in study area.

References

Gorsevski P. V., Brown M. K., Panter K., Onasch C. M., Simic A. and Snyder J., 2016. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: A case study in the Cuyahoga Valley National Park, Ohio. *Landslides* 13, 467-484.

Chang K. T., Liu J. K., Chang Y. M. and Kao C. S., 2010. An Accuracy Comparison for the Landslide Inventory with the BPNN and SVM Methods. *Gi4DM* 2010, Turino, Italy.

Chang K. T. and Liu J. K., 2004. Landslide features interpreted by neural network method using a high-resolution satellite image and digital topographic data. *Proceedings of the XXth ISPRS Congress*, Istanbul, Turkey.

Chen H., Zeng Z. and Tang, H., 2015 Landslide deformation prediction based on recurrent neural network. *Neural Processing Letters* 41, 169-178.

Dieu Tien Bui, Quoc Phi Nguyen, Nhat-Duc Hoang, Harald Klempe, 2016. A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. *Landslides* 14, 1-17.

Dieu Tien Bui, Binh Thai Pham, Quoc Phi Nguyen, Nhat Duc Hoang, 2016. Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of least squares support vector machines and differential evolution optimization: A case study in central Vietnam. *International Journal of Digital Earth* 9, 1077-1097.

Donahue J., Hendricks L. A., Guadarrama S., Rohrbach M., Venugopalan S., Saenko K., Darrell T., 2015. Long-term recurrent convolutional networks for visual recognition and description. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2625-2634.

Hochreiter S. and Schmidhuber J., 1997. Long short-term memory. *Neural Computation* 9, 1735-1780.

Lee H. and Park Y., 1991. Nonlinear system identification using recurrent networks. *Proceedings of the 1991 IEEE international joint conference on neural networks*, Vancouver, 2410-2415.

Quoc Phi Nguyen, Du Duong Bui, SangGi Hwang, Khac Uan Do, Thi Hoa Nguyen, 2018. Rainfall-triggered landslide and debris flow hazard assessment using data mining techniques: A comparison of Decision Trees, Artificial Neural Network and Support Vector Machines. *Proceedings of the 2018 Vietnam Water Cooperation Initiative (VACI 2018) - Highlights*. Science and Technics Publishing House, Hanoi, Vietnam, 138-141.

Nguyen Quoc Phi, Nguyen Kim Long, Nguyen Phuong, 2016. Natural hazard assessment using big data algorithms. *Workshop on Environmental risk assessment of mining sites in Central Vietnam using statistical models*. Hanoi, 28-45.

Yu W., 2004. Nonlinear system identification using discrete-time recurrent neural networks with stable learning algorithms. *Information Sciences* 158,131-147.