A Modelling Tool for Rainfall-triggered Landslide Susceptibility Mapping and Hazard Warning based on GIS and Machine Learning

Haiwei Zhou¹, Jianjun Yu², Hangjian Feng³* and Jie Huang⁴

¹ Hangzhou Real Estate Development Research Center, Hangzhou, Zhejiang, 310006, China
² Industry Solutions Research and Development Center, Schneider Electric Software Holdings, Singapore, 486057, Singapore
³ Zhejiang Institute of Geology and Mineral Resources, Hangzhou, Zhejiang, 310007, China
⁴ School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China

*Corresponding author’s e-mail: pcerma@foxmail.com

Abstract. The landslide susceptibility mapping and hazard warning are widely adopted tools by the government, stakeholders and the public for landslide disaster preparedness and emergency planning. This study presented a modelling tool based on geographic information system (GIS) and machine learning to aid the two-step modelling procedure. The machine learning methods including artificial neural networks, support vector machines, and logistic regression were integrated into the GIS environment for modelling landslide susceptibility to simplify and automate the routines of model training, verification and prediction. Then, the meta-element model was employed to take the landslide susceptibility, antecedent effective rainfall and 24-hour forecasted rainfall as inputs to determine the landslide hazard level. The architecture to deploy the established meta-element model for real-time landslide hazard warning was also proposed. A study case in Chunan, China was selected to demonstrate the applicability of the modelling tool to aid landslide susceptibility mapping and real-time hazard warning in response to a typhoon event. The developed modelling tool was desired to evolve into cloud computing architecture to facilitate easy-to-reuse and uplift its scalability.

1. Introduction

The landslides triggered by heavy or prolonged rainfalls are the uppermost destructive natural disasters in China and across the world. They have resulted in tremendous life and economic losses, asset damages, and infrastructure interruptions etc. The landslide susceptibility mapping and hazard warning system are widely adopted tools for landslide disaster preparedness and mitigation. These are especially important for developing countries considering the increasing exposure and vulnerability due to the rapid urbanization and demographic growth. The recent data explosion and technology advancement have shed the lights for smart disaster management, whilst challenged the pathway of contemporary technology adoption. In the context of landslide, it is desired to have a modelling tool to compile the data from various sources and be empowered with analytical capabilities to aid best
practices in landslide susceptibility mapping and hazard warning; hence, to facilitate stakeholders plan and prepare for the emergency responses.

Landslide susceptibility mapping, as a fundamental scientific tool for supporting engineering design and planning, indicates the proneness of the ground to landslides [1, 2]. Numerous studies of landslide susceptibility mapping have been reported ranging from methodology development to case studies [3, 4, 5]. They all share a similar modelling process, which firstly relates terrain, geologic, geomorphologic and environmental factors to historical landslides, and then project the relationship to the study area to produce the susceptibility map. However, tremendous efforts are expected in developing landslide susceptibility models through repeated trial-and-error in order to obtain a scientific sound map. This requires a modelling tool to help engineers avoid the tedious and repeated data process tasks in such a modelling procedure, whereas emphasis on examining the scientific validity.

The landslide hazard warning model takes the landslide susceptibility and forecasted rainfall as inputs to determine the hazard level at geographical locations through predefined rules and issue the warning if necessary. A plenty of hazard warning systems [6] have been in operation in North American (e.g. Canada and USA) [7], European (e.g. Italy and Spain) [8, 9] and Asian countries (e.g. China and Japan) [10, 11]. Considering the hazard warning accuracy is crucial to effective emergency responses, the modelling tool is desired to optimize the landslide hazard warning model development, validation and deployment.

Furthermore, the geographic information system (GIS) has been an integrated user-interactive platform for disaster risk management [12, 13, 14, 15] due to its capability to store, manipulate, analyse and visualize spatial and temporal data. How to integrate the modern machine learning techniques into such a GIS environment to empower the modelling tool with spatial analytic and prediction capacities is of particular importance.

Therefore, the objective of this study is to propose a two-step workflow for landslide susceptibility mapping and hazard warning system setup. A GIS based modelling tool is developed to aid the modelling procedure, where the machine learning techniques including artificial networks (ANN), support vector machines (SVN) and logistic regression (LR) are integrated for spatial prediction, and the meta-element model is employed for hazard level determination. A system architecture of integrating the tools and models for establishing the real-time forecasting software services is proposed. A real-world study case in Chunan, Hangzhou, China is selected for demonstration.

2. Modelling workflow

Figure 1 illustrates the two-step workflow for modelling landslide susceptibility and establishing the hazard warning system. There are three categories of data sources, including: (i) a point layer representing the meteorological stations, which links to the collected rainfall time series; (ii) landslide inventories recording the location of historical landslides with relevant characteristics such as date, magnitude, and size; and (iii) a group of geo-environment layers (i.e. digital elevation model, lithology and land use etc.) and social economic layers (i.e. settlements and infrastructures).

In step 1, the machine learning models are employed to model the landslide susceptibility. Firstly, the study area is divided into grids with specified spatial resolution (e.g. 30 m). Secondly, these grids are overlaid with geo-environmental layers to extract the information of landslide and stable slope samples for model training and verification. The established machine learning models are finally used to infer the probability of landslide occurrence for all grids over the study area to produce the landslide susceptibility map. In step 2, the matter-element model [16] is used to build up real-time landslide hazard warning system. The hazard warning model takes the landslide susceptibility, antecedent effective rainfall, and 24-hour ahead forecasted rainfall as inputs. The rules for hazard level classification is based on analysis of historical landslides in relation to rainfall patterns. Both the landslide susceptibility map and the predicted hazard warning map can be projected to social economic layers for vulnerability analysis and emergency planning.
2.1. Landslide susceptibility mapping based on machine learning

A desktop-based modelling tool was developed based on ArcGIS platform in Microsoft .Net (i.e. C# programming language) development environment. By encapsulating the ArcGIS Engine components, the GIS functionalities such as data import/export, visualization, complex query and spatial analysis could be implemented readily to manage and manipulate spatial and temporal data. The machine learning library AForge.Net was integrated into the GIS platform to automate the process of landslide susceptibility model training and verification. Powered by GIS, the required inputs of training samples could be easily extracted from various geo-environmental layers to feed the machine learning model. The seamless integration of GIS and machine learning would allow the established model to be applied to spatial prediction easily to aid landslide susceptibility mapping.

Considering the advantages and limitations that each machine learning method possesses, we selected ANN, SVM, and LR as the options for modelling landslide susceptibility. The ANN has the capability of self-learning and is easy-to-implement. The ANN model can be tuned flexibly through various combinations of network structure, number of hidden layers and neurons, transfer functions and learning algorithms etc. The SVN solves the regression problem based on a best hyperplane in a high-dimensional space through maximizing the functional margin. The Gaussian kernel function is introduced to obtain a nonlinear classifier with the purpose of achieving better prediction accuracy. However, the complex and black box modelling procedure of ANN and SVM limits the potential to examine the regression equations in order to uncover the correlation between the geo-environmental factors and landslide susceptibility. On the contrast, LR approximates the relationship between independent variables (i.e. geo-environmental factors) and binary output (i.e. probability of landslide occurrence) with the following explicit equations:
\[ y = \frac{1}{1 + e^{-u}} \]  
\[ u = \sum_{i=1}^{n} b_i x_i + b_0 \]

where \( x_i \) is the value of the \( i \)th geo-environmental factor (totally \( n \)) at concerned grid; \( b_i \) is regression coefficient under estimation, which can help evaluate the contribution of each factor to landslides; \( y \) represents the probability of landslide with the value of 0 standing for stable slopes and that of 1 for its occurrence. LR shows its advantage of efficient parameter estimation and better performance in classifying binary data, hence, is favourable in landslide susceptibility modelling.

Figure 2. Illustration of landslide susceptibility mapping aided by the developed modelling tool

Figure 2 illustrates the common process of applying machine learning for landslide susceptibility mapping. The inferred susceptibility value is in a range from 0 to 1. The closer the value is to 1, the more plausible the landslide is to occur. Once the susceptibility of all grids over the study area is inferred, the calculated receiver operating characteristic curve (ROC) and the value of area under the curve (AUC) are calculated to verify the modelling accuracy. The developed tool simplifies the modelling procedure by automating a series of compulsory tasks such as data clean, model configuration, training and verification, which are time consuming but crucial in applying machine learning to solve practical problems. Thus, it will allow engineers to emphasize on the exploration of
best modelling practices in landslide susceptibility mapping and estimate the associated uncertainties to understand the risk.

2.2. Landslide hazard warning

The matter-element model was widely employed as a mean for landslide hazard warning in China. It is expressed as:

$$R_i = \left( \begin{array}{c} N_i \\ c_{i1} \\ c_{i2} \\ c_{i3} \end{array} \right) \left( \begin{array}{c} (a_{i1}, b_{i1}) \\ (a_{i2}, b_{i2}) \\ (a_{i3}, b_{i3}) \end{array} \right) (i = 1, 2, 3, 4)$$ (3)

where $R$ is the ordered triad (named as matter-element) to describe things in real world; $N$ represents the predefined 4 grades of hazard levels (i.e. matter) denoted as “high”, “moderate”, “low”, and “very low”; each hazard level consists of three elements to describe it, including antecedent effective rainfall ($c_{i1}$), 24-hour ahead forecasted rainfall ($c_{i2}$), and degree of landslide susceptibility ($c_{i3}$). A variable range quantified by minimum ($a_i$) and maximum ($b_i$) threshold is used to measure the corresponding element.

The degree of matter-element $R_i$ at each grid depending on the predefined hazard level ($R_i$) could be calculated as:

$$K_i = \sum_{j=1}^{n} \alpha_j K_{ij}$$ (4)

$$K_{ij}(x) = \begin{cases} -\rho(x), & a < x < b \\ \rho(x), & x \leq a \text{ or } x \geq b \end{cases}$$ (5)

$$\rho(x) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2}$$ (6)

$$\eta(x) = \left| x - \frac{c+d}{2} \right| - \frac{d-c}{2}$$ (7)

where $K_i$ is the dependent degree which is the weighted ($\alpha_j$) sum of the individual dependent degree ($K_{ij}$) of $j$th element; $x$ is the given value in $R_i$ representing a specific condition under evaluation; $a$ and $b$ are minimum and maximum value of variable ranges for a specific element; and accordingly, $c$ and $d$ define an overall variable range covering all of those matters. The $R_i$ is finally classified to the matter (i.e. hazard level) whose $K_i$ is largest among the others based on the assumption that a larger $K_i$ means a higher dependent degree between $R_i$ and $R_i$.

Based on historical rainfall and landslide inventories, the modelling tool determines the thresholds in meta-element model to enable the predicted hazard map achieving maximum AUC. The established matter-element model will be deployed as a software service for real-time landslide hazard level forecasting. As shown in figure 3, the overall architecture consists of: (i) a shared central data repository serving for all the software components in the system for various data storage and access; (ii) a remote data monitoring sub system to automatically collect the sensed and forecasted data (e.g. rainfall). The data transfer between the sub system and the central repository is based on general packet radio service (GPRS) or intranet; (iii) an operational forecasting service, a routine running on the server, to predict the real-time hazard level whenever the forecasted rainfall is available. For easier management purpose, the warning level is issued for each administrative village and published on a web-based data portal. The warning level is determined based on spatial overlay of resident distributions and predicted landslide hazard map; and (iv) a desktop-based modelling tool for interactive data visualization, model buildup, vulnerability analysis and mapping. In addition, the onsite landslide monitoring system is also possible to be integrated into the proposed framework to enhance accurate early warning for critical surveillance regions.
3. A case study in Chunan, China

A study case in Chunan county (figure 4a), Hangzhou city, China was used to demonstrate the applicability of the modelling tool for landslide susceptibility mapping and hazard warning. Chunan has an area of 4,427 km$^2$, covering a water body (i.e. Xinanjiang reservoir) of 537 km$^2$. According to recordings, the monthly mean temperature ranges from 5.2 °C in winter to 28.6 °C in summer. The amount of annual precipitation varies from 1,025 to 2,111 mm with an average of 1,478 mm. The high-intensity storm is the primary triggering factor of the landslides, which has resulted in totally 596 landslides over the period from 1990 to 2013. The storms and landslides show a positive correlation, occurring majorly in June and July simultaneously.

As an example, the ANN model was selected for landslide susceptibility mapping. Aided by the modelling tool, totally 451 (i.e. 75%) out of 596 landslides are randomly selected as positive samples; meanwhile the same number of negative samples are selected in spatial randomly from stable slopes for model training. The values of geo-environmental factors including elevation, slope, aspect, curvature, lithology, distances to drainage, fault, road, and land use were extracted for concerned grids to feed the ANN model. Through trial-and-error, the feedforward ANN was constructed with 3 layers, including an input layer with 9 inputs, an output layer and a hidden layer settled with 5 neurons. The model was trained using Levenberg-Marquardt algorithm and verified using the rest 145 landslide samples (i.e. 25%). The trained ANN model achieved the AUC of 85.7% and projected to the study area to produce the landslide susceptibility map (figure 4b).

Table 1. showed the derived thresholds for defining the matter-element model based on analysis of historical landslides in relation to the heavy downpours. In the event of 2014 Typhoon Matmo, the 24-hour rainfall exceeding 100 mm mainly fell in southeast, northwest and western Chunan. Figure 4c showed the landslide hazard warning map issued at administrative village level. Accordingly, response actions for specific hazard levels could be planned. Practically, for the villages issued with moderate
hazard level, the inspection and surveillance of the hazard-proven areas was executed, and the evacuation plan was implemented.

| Table 1. Definition of matter-element model for landslide hazard warning. |
|--------------------------------|-----------------|-----------------|-----------------|
| Hazard level  | Landslide susceptibility | Forecasted rainfall (mm) | Antecedent effective rainfall (mm) |
| High       | (0.81, 1]  | >150            | >250            |
| Moderate   | (0.64, 0.81] | (90, 150] | (125, 250] |
| Low        | (0.46, 0.64] | (50, 90] | (50, 125] |
| Very low   | [0, 0.46]  | [0, 50]        | [0, 50]        |

Figure 4 (a) study case of Chunan, China; (b) produced landslide susceptibility map using artificial neural networks; (c) issued landslide hazard warning map for 2014 Typhoon Matmo

4. Conclusions
This study presented a GIS and machine learning based modelling tool to aid landslide susceptibility mapping and hazard warning. A two-step workflow was proposed, including: (i) integrating artificial neural networks, support vector machines and logistic regression in a GIS environment for modelling susceptibility; and (ii) employment of meta-element model for landslide hazard level determination and its deployment for real-time forecasting. A case study in Chunan was selected to demonstrate the application of the modelling tool to aid effective emergency responses in a Typhoon event. The design and development of the modelling tool was aimed for standalone users only. Hence, it is desired to evolve into cloud computing architecture to gain wider adoption.

References
[1] Van Westen, C.J., Castellanos, E., Kuriakose, S.L. (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Eng. Geol., 102(3-4):112-131.
[2] Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K. (2012) Landslide inventory maps: New tools for an old problem. Earth-Sci. Rev., 112(1-2):42-66.
[3] Xu, C., Xu, X., Dai, F., Saraf, A.K. (2012) Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. Comput. and Geosci., 46:317-329.
[4] Feng, H.J., Yu, J.J., Zheng, J.L., Tang, X.M., Peng, C.C. (2016) Evaluation of different models in rainfall-triggered landslide susceptibility mapping: a case study in Chunan, southeast China. Environ. Earth Sci., 75:1399.
[5] Lin, L., Lin, Q.G., Wang, Y. (2017) Landslide susceptibility mapping on a global scale using the method of logistic regression. Nat. Hazards Earth Syst. Sci., 17:1411-1424.
[6] Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M., Melillo, M. (2020) Geographical landslide early warning systems. Earth-Sci. Rev., 200:102973.

[7] Baum, R.L., Godt, J.W. (2010) Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides, 7:259-272.

[8] Berenguer, M., Sempere-Torres, D., Hürlimann, M. (2015) Debris-flow forecasting at regional scale by combining susceptibility mapping and radar rainfall. Nat. Hazards Earth Syst. Sci., 15(3):587-602.

[9] Segoni, S., Tofani, V., Rosi, A., Catani, F., Casagli, N. (2018) Combination of Rainfall Thresholds and Susceptibility Maps for Dynamic Landslide Hazard Assessment at Regional Scale. Front. Earth Sci., https://doi.org/10.3389/feart.2018.00085

[10] Osanai N, Shimizu T, Kuramoto K, Kojima S, Noro T (2010) Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network. Landslides, 7:325–338.

[11] Wu, Y., Chen, L., Cheng, C., Yin, K., Török, Á. (2014) GIS-based landslide hazard predicting system and its real-time test during a typhoon, Zhejiang Province, Southeast China. Eng. Geol., 175:9-21.

[12] Li, C., Ma, T., Zhu, X. (2010) aiNet- and GIS-based regional prediction system for the spatial and temporal probability of rainfall-triggered landslides. Nat. Hazards, 52:57-78.

[13] Yu. J.J., Qin, X.S., Larsen, L.C., Larsen, O., Jayasooriya, A., Shen, X.L. (2012) A GIS-based management and publication framework for data handling of numerical model results. Adv. Eng. Softw., 45(1):360-369.

[14] Bragagnolo, L., da Silva, R.V., Grzybowski, J.M.V. (2020) Landslide susceptibility mapping with r.landslide: A free open-source GIS-integrated tool based on Artificial Neural Networks. Environ. Model. Softw., 123:104565.

[15] Sahin, E.K., Colkesen, I., Acmali, S.S., Akgun, A., Aydinoglu, A.C. (2020) Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. Comput. and Geosci., 144:104592.

[16] Cai, W. (1999) Extension theory and its application. Chinese Sci. Bull., 44(17):1538-1548.