Landslide Susceptibility Mapping using Genetic Expression Programming

Maryamsadat Hosseini¹,  *, Rouzbeh Shad², Samsung Lim¹

¹ School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia
² Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Maryamsadat.hosseini@unsw.edu.au

Abstract. The increasing demand for land use and the mismanagement of lands have caused the increase of landslides around the world. It is important to recognize the landslide characteristics and the determining factors that influence this phenomenon in order to mitigate the adverse economic and environmental impacts. This study aims to estimate the landslide susceptibility in an area of Siahkal at Gilan province, Iran, by formulating a model using Gene expression programming (GEP). Seven condition factors including altitude, aspect, slope, proximity to rivers, proximity to faults, land use and lithology, were used in this research. The proposed model was developed as an equation, and its accuracy was assessed by the area under an Receiver Operating Characteristic curve that shows 0.82 and 0.77 for the training data and the test data, respectively. The result of this research was also evaluated with the inventory map over the study area which was constructed by field surveying and interpretation of airborne/satellite images. Our landslide susceptibility map indicates that the northern part of the area has the highest possibility for landslide which is in agreement with the landslide inventory map of the previous landslide.

1. Introduction

Natural hazards can cause social and economic crises and might lead to numerous casualties [1; 2]. Landslides are one of the devastating geological hazards which may cause serious damages to infrastructures and environments [1; 3]. Therefore, it is crucial to predict and determine landslide-prone areas in order to avoid the adverse effects of landslides on people’s lives and properties [1; 4]. Different methods have been used to understand the importance of various factors that are affecting landslide susceptibility mapping [4]. These methods include not only statistical methods but also machine learning methods [5]. The necessity of normal distributions of environmental factors for statistical models is one of the drawbacks of the statistical methods which are inherently linear [5]. On the other hand, advances in computing technologies have provided an opportunity to further statistical methods into machine learning methods [6]. Machine learning techniques can use input variables without statistical regularity [5]. Well-known machine learning methods include Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR) and Artificial Neural Networks (ANN) [5-7].

ANN was applied to landslide susceptibility mapping in several studies [8-12], while SVM [8; 9; 13; 14] and LR were also used for determining landslide susceptibility mapping [9; 12; 15; 16; 17].
has been used for complex modelling. However, it is one of the black-box models, and the model outcome from this method is challenging to get interpreted. The models generated from ANN have local optimum and overfitting limitations [5; 18; 19]. The SVM training process takes too much time, which is not an interpretable model [5; 13]. This encouraged researchers to use Evolutionary Algorithms (EA), namely Genetic Algorithm (GA), Genetic Programming (GP) and Genetic Expression Programming (GEP) [19; 20]. GP is less likely to trap in local optimums, and it is capable of modelling a nonlinear problem such as landslide susceptibility [19; 21]. GP can define the model in the form of definite comprehensible equations that lead researchers to conduct studies based on evolutionary computation approaches [19].

Some of previous studies have used GEP to create a model based on different environmental factors that affect the landslide occurrences. In this research, GEP is also used to produce the susceptibility maps for landslides. GEP was applied on seven conditional factors including altitude, aspect, slope, proximity to rivers, proximity to faults, land use and lithology. Firstly, the conditional factors were classified, and then GEP was used to determine the relations between different classes. Finally, the generated equation from GEP from all of the condition factors was used to create landslide susceptibility maps.

2. Methods

2.1. Study area

Our study area, Siahkal, is located between latitudes 36°52’30” N and 37°00’00” N, and longitudes 49°52’30” E and 50°00’00” E, with an area of around 70 km² in Gilan province (North of Iran) (Figure 1A and 1B). The region consists of highlands with the lowest level of nearly 300 m and the highest surface of about 2,500 m above sea level. A substantial amount of the study area is covered by rainforests followed by farmlands being the second greatest terrain feature. The climate in the region is cold and humid. Moreover, due to a high level of evaporation caused by the Caspian Sea, there is significant precipitation in the study area. High humidity in the area results in slight temperature fluctuations between summer and winter. In the presence of features such as humidity, frequent precipitations and mountainous terrain, the area holds a critically high potential for landslides.

![Figure 1. Location of the study area. A) Map of Iran, B) location of study area in Gilan provience and C) landslide inventory map in case study area.](image-url)
2.2. Conditioning factors

We chose different condition factors to determine the landslide susceptibility on the basis of landslide types, data availability for our selected area and previous research [1; 2; 4-6; 22-28]. We used the aforementioned seven condition factors which greatly influence the chances of landslides. In order to collect and prepare data, a topographical map (1:25,000 scale) of the region was obtained from the National Cartographic Center (NCC) of Iran along with a geological map (1:100,000 scale) from the Geological Survey of Iran (GSI).

Landslide inventory maps were used to train and test the GEP-generated model (Figure 1C). The inventory maps (1:50,000 scale) were gathered from the Forests, Range and Watershed Management Organization (FRWO) of Iran. The inventory map used in this research was generated from field surveying, aerial photographs and satellite images. The seven conditioning factors were classified using ArcGIS. The elevation data was categorized into ten classes (Figure 2A). The aspect data was classified into eight directional classes plus one directionless class, namely northward, north-eastward, eastward, south-eastward, southward, south-westward, westward, north-westward and flat (Figure 2B). The slope data was divided into six classes with equal intervals (Figure 2C). Euclidean distance was used to calculate the proximity to rivers and faults (Figure 2D and 2E, respectively). Distance from rivers was classified into five classes with equal intervals, and distance from faults was categorized into four classes (0-100, 100-200, 200-500, >500). The lithology data has 13 types of rocks which are present in the study area (Figure 2F). The land use data has three classes, namely garden, cultivation and forest (Figure 2G).

Figure 2. Conditioning factors maps A) elevation, B) aspect, C) slope and D) distance from river. The rest of the maps are presented in the continued Figure.
Figure 2. continued. Conditioning factors maps E) distance from fault, F) lithology, G) land use.

2.3. Gene Expression Programming (GEP)
GEP is one of Evolutionary Algorithms (EA) that acts like biological evolution that the fittest individuals survive [7; 29]. This algorithm was first introduced in 2001 [30]. The package used for landslide modelling in this research was GeneXproTools. The GEP algorithm begins with creating chromosomes of the initial population [30]. A set of functions were defined as a character for the encoding of the chromos (Table 1). Then the decision boundaries were defined to evaluate the fitness of the individuals [7]. The evaluation of model fitness was conducted by the area under the Receiver Operating Characteristic (ROC) curve. Selection programs will come to effect when the termination criteria are not satisfied. Then, several genetic operations, such as replication, mutations, etc. were applied to reproduce the next generations. In the replication step, two selected parents exchanged their genetic elements, and in mutation, certain gene elements will be modified randomly [7].

Table 1. Functions used in GEP [31].

| Name                | Notation | Definition   |
|---------------------|----------|--------------|
| Addition            | +        | (x+y)        |
| Subtraction         | -        | (x-y)        |
| Multiplication      | *        | (x*y)        |
| Division            | /        | (x/y)        |
| Square root         | Sqrt     | sqrt(x)      |
| Exponential         | Exp      | exp(x)       |
| Natural logarithm   | Ln       | ln(x)        |
| Logarithm of base 10| Log      | log(x)       |
| X to the power      | Xn       | x^n          |
| Arctangent          | Atan     | arctan(x)    |
3. Results and Discussion

To model the landslide phenomenon, this study uses GEP due to several advantages including having chromosomes with simple interconnected linear structures that are relatively small and convenient for genetic modification (e.g., repetition, mutation, recombination, rearrangement, etc.), as well as the ability to form tree subordinates in a variety of size and shape. The software used for this is GeneXproTools.

As for landslide susceptibility mapping, seven condition factors i.e., elevation, aspect, slope, proximity to rivers, proximity to faults, lithology and land use, were initially prepared in ArcGIS and then passed as an input to GeneXproTools. The data were divided into a training data (70%) and a test data (30%), for training and testing the model, respectively. The area under the ROC curve (AUC) was used as a performance measure. AUCs for the training and test data were 0.82 and 0.77, respectively (Figure 3). An AUC between 0.7-0.8 was considered as an acceptable and >0.8 as good classification performance [7]. This means that the obtained results in our research are reliable.

The best output model, considering an efficient program structure in this study, is provided in Equation (1):

\[
d_3 \cdot \tan^{-1}(d_7 \cdot \tan^{-1}(\sqrt{c \cdot d_6})) + d_6 \cdot d_4 \cdot (\sqrt{(d_4 \cdot d_3)} + d_3)^2 + \sqrt{\left(\frac{d_5}{e(d_3-d_2)}\right) \cdot d_3 + (d_6 \cdot \sqrt{\tan^{-1}(d_7 \cdot (\tan^{-1}d_2)^2)})}
\]

where \(d_0, d_1, ..., d_6\) represent aspect, distance to faults, distance to rivers, lithology, slope, elevation, land use (cultivation), land use (forest), land use (gardens), respectively. The value for \(c\) in the equation was 2.003.

![Figure 3. AUC representing the model qualification.](image)

Finally, a landslide susceptibility map of the study area was generated by using the proposed model (Figure 4). The landslide susceptibility was classified from low to high by the natural break method in ArcGIS. The result indicates that the majority of the landslide-prone regions are identified in the north and north-west of the study area. Similarly, the inventory map indicates that the northern region had the biggest landslide in the past. This indicates that our model can predict landslides accurately.
Figure 4. Landslide susceptibility map produced from GEP method. The map overlayed on the map of the study area using Google Earth.

4. Conclusions
The adverse impacts of landslides encourage researchers to develop models for predicting this phenomenon in order to help decision-makers to mitigate the effects of landslide hazards. Landslides are a complicated nonlinear phenomenon, and we proved that GEP is an appropriate algorithm to address the problem because it deals with complex nonlinear problems. Also, unlike common approaches such as artificial neural networks which yield vague solutions for such issues, GEP provides results in the form of a clear and readily understandable equation. GEP was less likely to be trapped in local optimum compared to other evolutionary counterparts such as GA and GP. Therefore, to map potential landslide areas, this approach is strongly recommended due to its accuracy, simplicity, and ability to provide definite formula. The assessment on the basis of AUC shows that the generated model for our study area is reliable as evidenced by 0.82 and 0.77 for the training data and test data, respectively. The landslide susceptibility map generated from the proposed model is also in agreement with the reference inventory map.

References
[1] Pham B T, Nguyen-Thoi T, Qi C, Phong T Van, Dou J, Ho L S, Le H Van and Prakash I (2020). Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping. Catena, 195.
[2] Wang Y, Fang Z, and Hong H (2019). Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the Total Environment, 666, 975–993. https://doi.org/10.1016/j.scitotenv.2019.02.263.
[3] Kadavi P R, Lee C W, and Lee S (2018). Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sensing, 10. https://doi.org/10.3390/rs10081252.
[4] Lee S, Hong S M, and Jung H S (2017). A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. Sustainability (Switzerland), 9. https://doi.org/10.3390/su9010048.
[5] Huang F, Yin, K, Huang J, Gui L, and Wang P (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. Engineering Geology, 223, 11–22. https://doi.org/10.1016/j.enggeo.2017.04.013.

[6] Hong H, Liu, J, Bui D T, Pradhan B, Acharya T D, Pham B T, Zhu A X, Chen W, and Ahmad, B Bin (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, 399–413. https://doi.org/10.1016/j.catena.2018.01.005.

[7] Hoang N D, and Tien Bui D (2018). Spatial prediction of rainfall-induced shallow landslides using gene expression programming integrated with GIS: a case study in Vietnam. Natural Hazards, 92, 1871–1887. https://doi.org/10.1007/s11069-018-3286-z.

[8] Chen W, Pourghasemi H R, Kornejady A, and Zhang N (2017). Landslide spatial modelling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma, 305, 314–327. https://doi.org/10.1016/j.geoderma.2017.06.020.

[9] Kalantar B, Pradhan B, Amir Naghibi S, Motevalli A, and Mansor S (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). Geomatics, Natural Hazards and Risk, 9, 49–69. https://doi.org/10.1080/17475705.2017.1407368.

[10] Kanungo D P, Arora M K, Sarkar S, and Gupta R P (2006). A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Engineering Geology, 85, 347–366. https://doi.org/10.1016/j.enggeo.2006.03.004.

[11] Kawabata D, and Bandibas J (2009). Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology, 113, 97–109. https://doi.org/10.1016/j.geomorph.2009.06.006.

[12] Park S, Choi C, Kim B, and Kim J (2013). Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environmental Earth Sciences, 68, 1443–1464. https://doi.org/10.1007/s12665-012-1842-5.

[13] Marjanović M, Kovacević M, Bajat B, and Voženílek V (2011). Landslide susceptibility assessment using SVM machine learning algorithm. Engineering Geology, 123, 225–234. https://doi.org/10.1016/j.enggeo.2011.09.006.

[14] Yao X, Tham L G, and Dai F C (2008). Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. Geomorphology, 101, 572–582. https://doi.org/10.1016/j.geomorph.2008.02.011.

[15] Erener A, Mutlu A, and Sebnem Düzgün H (2016). A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). Engineering Geology, 203, 45–55. https://doi.org/10.1016/j.enggeo.2015.09.007.

[16] Mandal S, and Mandal K (2018). Modelling and mapping landslide susceptibility zones using GIS based multivariate binary logistic regression (LR) model in the Rorachu river basin of eastern Sikkim Himalaya, India. Modeling Earth Systems and Environment, 4, 69–88. https://doi.org/10.1007/s40808-018-0426-0.

[17] Mondal S, and Mandal S (2018). RS and GIS-based landslide susceptibility mapping of the Balason River basin, Darjeeling Himalaya, using logistic regression (LR) model. Georisk, 12, 29–44. https://doi.org/10.1080/17499518.2017.1347949.

[18] Gong Y, Zhang Y, Lan S, and Wang H (2016). A Comparative Study of Artificial Neural Networks, Support Vector Machines and Adaptive Neuro Fuzzy Inference System for Forecasting Groundwater Levels near Lake Okeechobee, Florida. Water Resources Management, 30, 375–391. https://doi.org/10.1007/s11269-015-1167-8.
[19] Nourani V, Pradhan B, Ghaffari H, and Sharifi S S (2014). Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Natural Hazards, 71, 523–547. https://doi.org/10.1007/s11069-013-0932-3.

[20] Kavzoglu T, Kutlug Sahin E, and Colkesen I (2015). Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. Engineering Geology, 192, 101–112. https://doi.org/10.1016/j.enggeo.2015.04.004.

[21] Nourani V, Komasi M, and Alami M T (2012). Hybrid Wavelet–Genetic Programming Approach to Optimize ANN Modeling of Rainfall–Runoff Process. Journal of Hydrologic Engineering, 17, 724–741. https://doi.org/10.1061/(asce)he.1943-5584.0000506.

[22] Dikshit A, Sarkar R, Pradhan B, Jena R, Drukpa D, and Alamri A M (2020). Temporal probability assessment and its use in landslide susceptibility mapping for Eastern Bhutan. Water (Switzerland), 12. https://doi.org/10.3390/w12010267.

[23] Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, and Hübl J (2019). Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. Science of the Total Environment, 653, 801–814. https://doi.org/10.1016/j.scitotenv.2018.10.431.

[24] Moayedi H, Mehrabi M, Mosallanezhad M, Rashid A S A, and Pradhan, B (2019). Modification of landslide susceptibility mapping using optimized PSO-ANN technique. Engineering with Computers, 35, 967–984. https://doi.org/10.1007/s00366-018-0644-0.

[25] Pourghasemi H R, Kornejady A, Kerle N, and Shabani F (2020). Investigating the effects of different landslide positioning techniques, landslide partitioning approaches, and presence-absence balances on landslide susceptibility mapping. Catena, 187. https://doi.org/10.1016/j.catena.2019.104364.

[26] Sahin E K, Colkesen I, Acmali S S, Akgun A, and Aydinoglu A C (2020). Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. Computers and Geosciences, 144. https://doi.org/10.1016/j.cageo.2020.104592.

[27] Stanley T, and Kirschbaum D B (2017). A heuristic approach to global landslide susceptibility mapping. Natural Hazards, 87, 145–164. https://doi.org/10.1007/s11069-017-2757-y.

[28] Wu Y, Ke Y, Chen Z, Liang S, Zhao H, and Hong H (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena, 187. https://doi.org/10.1016/j.catena.2019.104396.

[29] Kayadelen C (2011). Soil liquefaction modelling by Genetic Expression Programming and Neuro-Fuzzy. Expert Systems with Applications, 38, 4080–4087. https://doi.org/10.1016/j.eswa.2010.09.071.

[30] Ferreira C (2001). Gene expression programming: a new adaptive algorithm for solving problems. ArXiv Preprint Cs/0102027.

[31] Zounemat-Kermani M, Rajaei T, Ramezani-Charmahineh A and Adamowski J F (2017). Estimating the aeration coefficient and air demand in bottom outlet conduits of dams using GEP and decision tree methods. Flow Measurement and Instrumentation, 54, 9-19. https://doi.org/10.1016/j.flowmeasinst.2016.11.004.