ΕΓΕ 2019
15th International Congress of the Geological Society of Greece
Athens, 22-24 May, 2019 | Harokopio University of Athens, Greece
Bulletin of the Geological Society of Greece, Sp. Pub. **7** Ext. Abs. GSG2019-205

Applying Machine learning and GIS in landslide susceptibility assessment. The case of Krikeliotis water basin, Evritania, Greece

A.Tsampa[1], P.Tsangaratos[2,3]

(1) School of Science & Technology, Hellenic Open University, Parodos Aristotelous 18, 26335 Patra, std115817@ac.eap.gr, tsampa.areti@gmail.com
(2) School of Science & Technology, Hellenic Open University, Parodos Aristotelous 18, 26335 Patra, tsaggaratos.paraskevas@ac.eap.gr
(3) School of Mining and Metallurgical Engineering, National Technical University of Athens, 9, Ir. Polytechniou str., 15780, Zografou, Athens, Greece, ptsag@metal.ntua.gr

Landslides are considered as one of the most disastrous natural hazards worldwide which are responsible for adverse affects on the natural and human environment. The non-linear and complex nature of the physical but also man-made processes which leads to the evolution of landslides makes the prediction of landslides a very difficult task (Pourghasemi *et al.*, 2018). Landslide susceptibility assessments are the main investigation tool capable of providing valuable information concerning the relation between landslides and landslide-related variables but also the main investigation tool for zoning landslide prone areas (Tien Bui *et al.*, 2017). Nowadays, it is common practice to apply Machine learning algorithms along with Geographic Information Systems (GIS) so as to develop landslide susceptibility models. Numerous applications are found in the literature involving fuzzy logic algorithms, artificial neural networks and neuro-fuzzy models, ensemble tree-based models and evolutionary population based algorithms (Hong *et al.*, 2017; Pham *et al.*, 2018; Pourghasemi *et al*., 2018; Chen *et al.*, 2019). In this context, the main purpose of the present study was to produce a predictive spatial model for landslide susceptibility by using Machine Learning methods and GIS. In particular, a Random Forest (RF) model (Dou *et al.*, 2019) was used as the base learning algorithm for the development of the predictive spatial model whereas the management of the landslide related variables were achieved using GIS. The developed methodology involved several phases. During the first phase nine spatial variables related to landslide phenomena were selected and the data of 85 landslide events were analyzed. Analytically, the elevation, the slope angle, slope aspect, the distance from the river network, the profile curvature, the plan curvature, the Topographic Wetness Index, the geological formations and the distance from the tectonic features were the nine variables identified as most important and relevant to the landslide phenomena recorded in the research area. The second phase involved classifying each variable and weighting them according to the results obtained by the method Weight of Evidence. The third phase involved a multi-collinearity analysis in order to identify the existence or not of correlations between the landslide-related variables and to decide their usage and the application of the Learning Vector Quantization (LVQ) algorithm, so as to evaluate the predictive ability of the variables. During the third phase, the initial database was separated into training (70% of the total number of incidence) and validating (the remaining 30%) subsets. The last phase involved the implementation of the RF model and the construction of the landslide susceptibility map, whereas to evaluate the performance of the developed methodology the area under the success and predictive curve (AUC) were used. The water basin of the Krikeliotis river located in the Municipality of Evritania, Greece was selected as a test site to evaluate the predictive performance of the developed methodology whereas several R packages were used, based on R a language and environment for statistical analysis and graphical presentation (R Core Team, 2017), whereas ArcGIS 10.3.1 (ESRI, 2013) was used for accessing the spatial data and generating the landslide susceptibility maps. Based on the results of the multi-collinearity analysis all variables were used during the analysis, since no serious collinerarity was detected, whereas the LVQ algorithm evaluated as the most important and critical variable the variable associated with the geological formations, followed by the altitude and the morphological gradient (Table 1).

Table 1. Results of multi-collinearity analysis and importance based on LVQ algorithm.

Landslide variables	**Variance Inflation Factor**	**Tolerance Index**	**Importance (LQV)**
Geology formations	1.719	0.581	0.75
Distance from tectonic features	1.146	0.872	0.59
Distance from river network	1.396	0.716	0.51
Elevation	1.391	0.719	0.74
Slope angle	1.752	0.570	0.73
Slope aspect	1.140	0.876	0.61
TWI	1.899	0.526	0.68
Plan curvature	1.515	0.660	0.53
Profile curvature	1.093	0.914	0.52

During the study a grid search technique was used so as to set the optimal structural parameters *mtry* (number of variables used in each model) and *ntree* (number of produced trees) used by the RF model. For the *mtry* parameter it was set to 7 and for the *ntree* parameter it was set to 2000. According to the results of the RF model, the most important variables based on the Mean Decrease Accuracy and the Mean Decrease Gini was the elevation and the geological formations. Figure 1 illustrates the produced by the RF model landslide susceptibility map in which the 21.36% of the study area is characterized as very high susceptible, the 24.90% as high susceptibility, while the 34.19% as low and very low landslide susceptibility. Furthermore, by analyzing the training and validation data, the RF model showed an AUC value of 0.8670 and 0.7851, respectively, whereas the relative landslide density for the very high susceptible zone was 0.8460 when using the training subset and 0.7242 when using the validation subset.

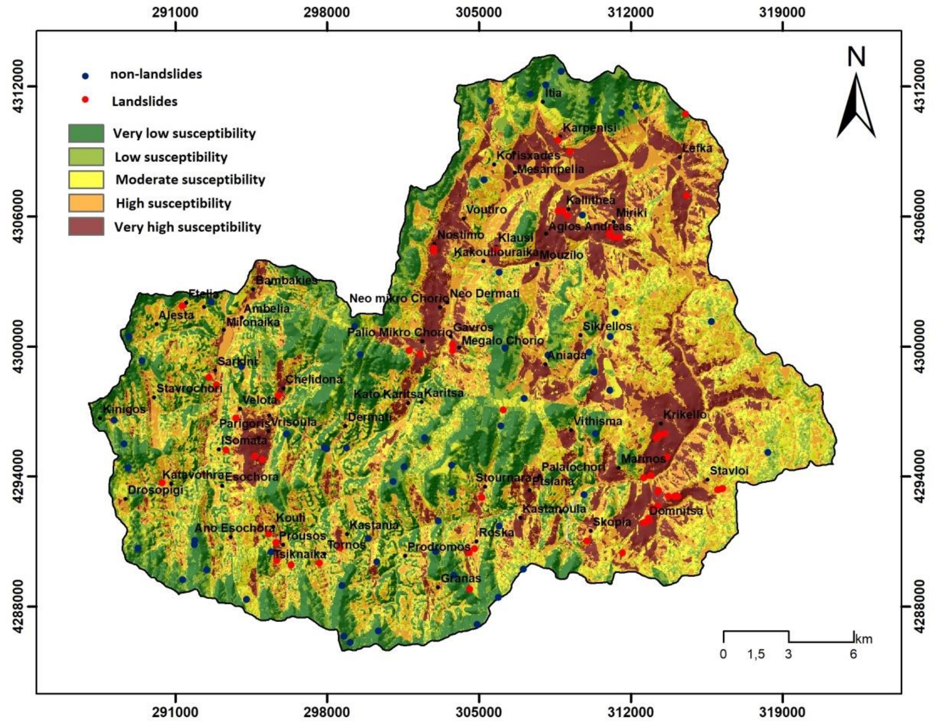

Figure 1. Landslide susceptibility based on RF model.

As a concluding remark one can highlight the fact that the identification of areas characterized by very high landslide susceptibility achieved by this study and also similar studies could be considered as a basic process that should precede any implementation of infrastructure development projects since it provides crucial information and knowledge concerning the mechanism responsible for landslide phenomena.

References

Chen, W., Panahi, M., Tsangaratos, P., Shahabi, H., Ilia, I., Panahi, S., Li, S., Jaafari, A., Ahmad B.B., 2019. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. Catena 172, 212-231.

Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Khosravi, K., Yang, Y., Pham, B. T., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci. Total Environ. 662, 332-346.

ESRI, 2013. ArcGIS Desktop: Release 10.1 Redlands, CA: Environmental Systems Research Institute.

Hong, H., Ilia, I., Tsangaratos, P., Chen, W., Xu, C., 2017. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. Geomorphology 290, 1–16.

Pham, B.T., Shirzadi, A., Tien Bui, D., Prakash, I., Dholakia, M.B., 2018. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation Forest for landslide susceptibility modeling: a case study in the Himalayan area, India. Int. J. Sediment Res. 33(2), 157-170.

Pourghasemi, H.R.,Yansari, Z.T., Panagos, P., Pradhan, B., 2018. Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). Arab. J. Geosci. 11, 193. https://doi.org/10.1007/s12517-018-3531-5.

R Core Team, 2017. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org.

Tien Bui, D., Tuan, T.A., Hoang, N.-D., Thanh, N.Q., Ngugen, D.B., Liem, N.V., Pradhan, B., 2017. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization, Landslides 14(2), 447-458.