Review on the Application of Machine Learning Methods in Landslide Susceptibility Mapping

Deborah Simon Mwakapesa[1], Ye Li[2], Wang Xiangtai[2], Guo Binbin[2] and Mao Yimin[2]
[1]*School of Civil and Surveying Engineering, Jiangxi University of Science and Technology, Ganzhou, China*
[2]*School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China*

Keywords: Machine Learning, Supervised Learning, Unsupervised Learning, Deep Learning, Landslide, Landslide Susceptibility Mapping

Abstract: Machine learning is a very important in computer science field which has gained attention in numerous applications. This paper reviewed various machine learning methods including supervised and unsupervised learning and highlighted their applications, advantages and disadvantages in landslide susceptibility mapping. The review has also mentioned the challenges of machine learning algorithms for achieving higher performance accuracy from the supervised and unsupervised learning algorithms during landslide susceptibility. Moreover, highlights on the application of deep learning methods as the current research in landslide susceptibility mapping has also been reported. Finally, this paper argued the necessity of thorough preparation of relevant and enough data being significant important to obtain high performance results from the review methods.

1 INTRODUCTION

Landslides (Cruden, 1991) involve the downward movement of various earth materials such as soil, mass or rocks, debris, and others, as a result of gravity. Population growth, settlement and economic development, as well as climatic changes are learnt to be the major triggers of landslides. These landslides have brought up highly variable impacts on both physical as well as human environment. It has been recorded that not less than 3.5 million kilometers square of the total land area in the world have been affected and are still susceptible to landslides (Dilley et al., 2005). Also, according to World Health Organization (WHO), between 1998 and 2017 approximately 4.8 million people were affected by landslides which also lead to more than 18,000 deaths. It is named among the very dangerous and disturbing disasters in the world. Thus, investigating and identifying areas susceptible to landslides for taking control as well as preventing measures is very important. One of the common measure is landslide susceptibility mapping (LSM, conducted on different landslide influencing factors such as geological, geomorphological and hydrological factors) using various methods such as Machine Learning (ML) which has gained much attention among researchers from different places in the world (Wang et al., 2020; Hu et al., 2020; Mao et al., 2021a; Mao et al., 2021b).

In the following sections, the machine learning methods are described as well as their applications in LSM are briefly highlighted.

2 REVIEW OF MACHINE LEARNING AND ITS APPLICATION IN LSM

2.1 Machine Learning (ML)

ML (Mitchell Tom, 1997) involves computer algorithms which improves through experience and by the use of data. ML algorithms construct models by learning from data and self-improve. These algorithms are applied in different applications including computer vision, medicine, speech recognition, and disaster prediction. ML is divided into four types: supervised learning, semi supervised learning, unsupervised learning and reinforcement learning. In LSM, the supervised and unsupervised

Mwakapesa, D., Li, Y., Xiangtai, W., Binbin, G. and Yimin, M.
Review on the Application of Machine Learning Methods in Landslide Susceptibility Mapping.
DOI: 10.5220/0010790400003167
In *Proceedings of the 1st International Conference on Innovation in Computer and Information Science (ICICIS 2021)*, pages 69-72
ISBN: 978-989-758-577-7

learning are commonly used (Buhmann, 1992; Boussemart et al., 2011).

2.1.1 Supervised Learning (SL)

In SL (Sathya and Annammna, 2013), methods such as classification, regression and prediction are trained using labeled examples, such as an input where the desired output (labels) is known. For example, having dataset labeled either landslide or non-landslide. The algorithm receives a set of inputs samples along with the corresponding label; the algorithm learns by comparing the actual output with the corresponding label to find errors and then adjusts the model accordingly. When there is an additional unlabeled data, the SL methods use patterns to predict the values of the label. Thus, in LSM, SL is applied in predicting future landslide events. Some SL algorithms that have commonly applied in LSM in recent years include support vector machine (SVM, Yu and Lu, 2018; Anik and Suli, 2020), logistic regression (Feby et al., 2020; Paul and Alejandra, 2021), classification and regression trees (Chen et al., 2017; Sun et al., 2021), decision trees (Mao et al., 2017; Kavzoglu et al., 2019; Guo et al., 2021) random forest (Chen et al., 2017; Sun et al., 2021), weight of evidence (Anik and Suli, 2020) and artificial neural network (Bragagnolo et al., 2020; Lucchese et al., 2021).

2.1.2 Unsupervised Learning (USL)

USL (Hinton and Sejnowski, 1999) such as clustering algorithms are used on data that has no labels, so the algorithm must determine where the data belong to with the aim of exploring the data and find its patterns. Up to date, there are very few USL algorithms that have been proposed in LSM, including k-means (Wang et al., 2017, Guo et al., 2021), Fuzzy C-means (FCM) and K-means particle swam optimization (KPSO, Wan Shiuan 2013; Wan Shiuan 2015), k-means and Hiearchical clustering (Pokharel et al., 2020), CA-AQD (Hu et al., 2020), AHC-OLID (Mao et al., 2021a), and OA-HD (Mao et al., 2021b). From the analysis of the current proposed USL methods in LSM, are hybrid methods, which are the modification of the traditional USL methods, while the traditional USL methods have not been explored in length as compared to the SL methods.

2.2 Discussion

In this paper, the study was conducted on the application of ML methods in landslide susceptibility mapping, on the basis of the developments that have been proposed and reported by researchers. The major application of ML methods in LSM can be observed in the area of SL algorithms as most of the studies published in various journals are based on SL. This is due to the advantages possessed by the SL methods, including: SL allows researchers to collect of produced data based on experience; this experience enhances performance criteria optimization; and SL gives an exact idea about the classes in the data such as landslide and non-landslide classes. Thus, these advantages make it easy to implement the SL methods in LSM. However, their applications are limited in various ways: inability to discover deep or unknown patterns in the data, thus, the results may not always be accurate; the accuracy of the methods depends on the available data, they require a lot of samples from the labels or classes for training to obtain high accuracy, whereby, in real situations, it is not easy to obtain landslide data especially when dealing with large study areas; also, the involved training process consumes a lot of computation time especially with large datasets from large study areas.

On the other side, the current proposed USL methods in LSM, have shown some advantages over the SL methods including: with USL, the methods learn and discover the features or patterns present in the data, then finds the similarities and dissimilarities in the data which make it easy to group them into different groups (classes) in absence of the data labels; discovering of features in the data make it easy to process the data even when other unlabeled data are added; also, this process does not consume a lot of computation time. Despite of their advantages they also have disadvantages, such as the in some situations, their results may not be very accurate as there is no training of data during the process in some cases, human intervention might be needed to validate the results; in LSM projects with real data, the USL involves feeding of data to the algorithm continuously which may result in inaccurate results as well as time consuming; also, when there are a lot of features in the data the process becomes complex.

However, from the above analysis on both cases, the performance of these methods depend on the available data. Thus, thorough and careful preparation of the data is a very significant stage in LSM while using these methods. Also, it has been observed that in both cases of ML methods, their ability to learn deep features from the data is very shallow, as they have one hidden layer or none. Thus, their performance results may not be very accurate when a data with deep and complex features is involved.

Furthermore, the LSM literature shows that currently, the research is directing to proposing LSM models based on deep learning methods (Goodfellow et al., 2016) which tend to have better features as compared to the former SL and USL proposed methods (Nhu et al., 2020). This is because the DL methods possess hidden layers or deep structures which facilitate the learning of deep and complex features in the data, thus the name Deep Learning. They also make it easy to process big datasets from larger study areas. So far, there are very studies that have been proposed and have so far shown better performance results compared to the prevailing methods. Some of the LSM deep learning models that have been proposed so far includes: deep neural networks (DNN, Kanu et al., 2021; Dong et al., 2020; Bui et al., 2020; Nhu et al., 2020; Dou et al., 2020); convolutional neural networks (Bui et al., 2020; Dou et al., 2020; Nhu et al., 2020; Yi et al., 2020; Fang et al., 2020; Ngo et al., 2021; Bragagnolo et al., 2021). And so far, these methods have shown promising performance results in their implementations. However, they also have some limitations, such as the fact that the DL models requires a lot of samples to train the models, and in cases where it is not easy to obtain many samples, the DL performance becomes limited.

3 CONCLUSIONS

This study was reviews the application of machine learning methods, the supervised and unsupervised learning, in landslide susceptibility mapping. The two types have been briefly discussed, they advantages and disadvantages have also been provided. At last, we also looked at the deep learning method which as per the literature review it has shown to perform better than the machine learning methods. This learning methodology has great significance. Although it have not been explored much as compared to machine learning, it can be very helpful in research. It has also been observed that, the performance of all the reviewed methods depends on the data. Therefore, the selection and preparation of relevant and enough data is crucial for the methods to work efficiently, especially with the deep learning. Moreover, this paper should also contribute to the collection of various machine learning application in LSM for easy reference.

ACKNOWLEDGEMENTS

This study was supported by the National Key Research and Development program (2018YDC1504705) and the National Natural Science Foundation of China (41562019).

REFERENCES

Cruden, M, 1991. A simple definition of a landslide. *Bulletin of the International Association of Engineering Geology*.

Buhmann, J, Kuhnel, H, 1992. Unsupervised and supervised data clustering with competitive neural networks. [Proceedings 1992] IJCNN International Joint Conference on Neural Networks.

Mitchell, T, 1997. Machine Learning. New York: McGraw Hill

Hinton, G, Sejnowski, T, 1999. Unsupervised Learning: Foundations of Neural Computation. *MIT Press*.

Dilley, M, Chen, S, Deichmann, U, Lerner, L, Arthur, L, Arnold, M, 2005. Natural Disaster Hotspots: *A Global Risk Analysis. Washington, DC: World Bank*.

Boussemart, Y, Mary, L, Cummings, Jonathan, F, Nicholas R, 2011. Supervised vs Unsupervised Learning for Operator State Modeling in Unmanned Vehicle Settings. *Journal of Aerospace Computing, Information, and Communication*.

Sathya, R, Annamma A, 2013. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*.

Wan, Shiuan, 2013. Entropy-based particle swarm optimization with clustering analysis on landslide susceptibility mapping. *Environmental Earth Science*.

Wan Shiuan, 2015. Construction of knowledge-based spatial decision support for landslide mapping using fuzzy clustering and KPSO analysis. *Arabian Journal of Geoscience*.

Goodfellow, I, Bengio, Y, Courville, A, 2016. Deep learning. *MIT Press*.

Chen W, Xiaoshen X, Jiale W, Biswajeet P, Haoyuan H, Dieu B, Zhao D, Jianquan M, 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility, *CATENA*.

Wang Q, Wang Y, Niu Q, 2017. Integration of information theory, k-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the three gorges area, China. *Remote Sensing*.

Kavzoglu T, Colkesen I, Sahin EK, 2018. Machine learning techniques in landslide susceptibility mapping: a survey and a case study. *Advances in Natural and Technological Hazards Research*.

Yu H, Lu Z, 2018. Review on landslide susceptibility mapping using support vector machines, *CATENA*.

Anik Saha, Sunil Saha, 2020. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling Himalaya, India. *Remote Sensing Applications: Society and Environment*.

Bo, Yu, Fang Chen, Chong Xu, 2020. Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Computers & Geosciences*.

Bragagnolo, L, R.V. da Silva, Grzybowski, J 2020. Landslide susceptibility mapping with r.landslide: A free open-source GIS-integrated tool based on Artificial Neural Networks. *Environmental Modelling & Software*.

Dong, D, Abolfazl, J, Mahmoud, B, Davood, M, Chongchong, Q, Hossein, M, Tran, P, Hai-Bang, L, Tien-Thinh, L, Phan, T, Chinh, L, Nguyen, Q, Bui, T, Binh Thai, P, 2020. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA*.

Bui, T, Paraskevas, T, Viet-Tien N, Ngo, L, Phan, T, 2020. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *CATENA*.

Feby B, Achu, L, Jimnisha, K, Ayisha A, Rajesh, R, 2020. Landslide susceptibility modelling using integrated evidential belief function based logistic regression method: A study from Southern Western Ghats, India. *Remote Sensing Applications: Society and Environment*.

Hu, J, Xu, K, Wang, G. et al., 2020. A novel landslide susceptibility mapping portrayed by OA-HD and K-medoids clustering algorithms. *Bulletin Engineering Geology and the Environment*.

Dou J, Ali P. Y, Abdelaziz M, Ataollah S, Hoang N, Yawar H, Ram A, Yulong C, Binh Thai P, Hiromitsu Y, 2020. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Science of the Total Environment*.

Pokharel B, Omar A, Ali A, et al., 2020. Spatial clustering and modelling for landslide susceptibility mapping in the north of the Kathmandu Valley, Nepal. *Landslides*.

Nhu V, Nhat-Duc H, Hieu N, Phuong Thao N, Tinh Thanh B, Pham H, Pijush S, Dieu Tien B, 2020. Effectiveness assessment of Keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area, *CATENA*.

Yi Y, Zhijie Z, Wanchang Z, Huihui J, Jianqiang Z, 2020. Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *CATENA*.

Fang Z, Yi W, Ling P, Haoyuan H, 2020. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers and Geosciences*.

Bragagnolo L; L.R, Rezende, R.V. da Silva, Grzybowski J.M.V, 2021. Convolutional neural networks applied to semantic segmentation of landslide scars. *CATENA*.

Guo Z, Yu S, Faming H, Xuanmei F, Jinsong H, 2021. Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geoscience Frontiers*.

Kanu M, Sunil S, Sujit M, 2021. Applying deep learning and benchmark machine learning algorithms for landslide susceptibility modelling in Rorachu river basin of Sikkim Himalaya, India. *Geoscience Frontiers*.

Lucchese L, Guilherme Garcia de Oliveira, Olavo Correa P, 2021. Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using Artificial Neural Networks. *CATENA*.

Mao Y, Deborah SM, Wang G, Yaser AN, Zhang M, 2021a. Landslide susceptibility modelling based on AHC-OLID clustering algorithm. *Advances in Space Research*.

Mao Y, Yican L, Deborah SM, Wang G, Yaser AN, Muhammad AK, Zhang M, 2021b. Innovative Landslide Susceptibility Mapping Portrayed by CA-AQD and K-Means Clustering Algorithms. *Advances in Civil Engineering*.

Ngo Thao T, Mahdi P, Khabat K, Omid G, Narges K, Artemi C, Saro L, 2021. Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. Geoscience Frontiers.

Paul Goyes-P, Alejandra Hernandez-R, 2021. Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in Popayan, Colombia. *Engineering Geology*.

Sun D, Jiahui Xu, Haijia W, Danzhou W, 2021. Assessment of landslide susceptibility mapping based on Bayesian hyper-parameter optimization: A comparison between logistic regression and random forest. *Engineering Geology*.