LANDSLIDE SUSCEPTIBILITY MAPPING USING XGBOOST MODEL IN CHITTAGONG DISTRICT, BANGLADESH

Conference Paper · January 2019

3 authors:

Torit Chakraborty
Khulna University

Md. Shaharier Alam
Asian Disaster Preparedness Center

Md Didarul Islam
Central Michigan University

LANDSLIDE SUSCEPTIBILITY MAPPING USING XGBOOST MODEL IN CHITTAGONG DISTRICT, BANGLADESH

Torit Chakraborty[1], Md Didarul Islam[2] and Md Shaharier alam[3]

ABSTRACT

Landslide is one of the most destructive geological hazards in the hilly areas of Bangladesh due to heavy monsoon rainfall within short period. Chittagong District area was selected as study area which experiences several landslides each year causing casualties and economic loss. In this study, two machine learning models including XGBoost and support vector machine, are applied for landslide susceptibility mapping based on eight distinct causative factors in GIS platform. The landslide location data are randomly divided into two groups: 70% for training and 30% for validation. The accuracy of the aforementioned models are compared among themselves to find out the most accurate model for landslide susceptibility mapping. The validation result shows that XGBoost model gains highest 93.63% kappa score, 88.23% AUC score and 94.53% accuracy score whereas SVM gains 89.77% kappa score, 84.38% AUC score and 91.02% accuracy score. As XGBoost model has relatively higher accuracy, it is used to generate the final landslide susceptibility mapping of the study area where the result shows that 32.47% areas are highly susceptible, 19.75% are moderately susceptible and 47.75% areas are low susceptible. This study is expected to be helpful in developing effective risk prevention and mitigation strategies.
Keyword: Landslide susceptibility mapping, XGBoost, Support Vector Machine(SVM), GIS

Introduction

Landslides are among the natural disasters that are often experienced in Bangladesh specially in Chittagong district due to its geological, geographical structure and short period torrential rain which are always resulting this hazard sometimes in a very destructive way causing severe economic loss. For that reason, it's very necessary to analysis the landslide risky zone to take effective planning intervention to minimize the loss.All the recent study shows that due to heavy rainfall for a short period of time, Chittagong district has experienced landslides every year so that landslide susceptibility mapping is very important to identify the risky zone.

Landslide susceptibility mapping can be done using both qualitative and quantitative approach. But in recent decades the use of quantitative approach has become very popular due to new development of geospatial technology and geo-computing methods which provide better accurate results. In this research, two very popular machine learning model(XGBoost and SVM) has used to perform the landslide susceptibility mapping. Different researcher had used different model to produce landslide susceptibility map such as logistic regression, multi linear regression (ahmed,2015),in cox's bazar, Bangladesh, SVM (huang,2018) in china, artificial neural network, SVM(zhuo,2018) in china. But in this paper, another machine learning model is used named XGBoost to produce the landslide susceptibility mapping in comparison to the frequently used SVM model. XGboost had been used by researchers(Dong,2018),(Georganos,2018) for land use/land cover (LULC) image classification. But here this model is used to produce the landslide susceptibility map. For both model, 8 causative factors are chosen such as elevation, soil type, rainfall, NDVI (normalized difference vegetation index), distance from river, land cover, flow accumulation, slope to find spatial correlation with the landslide location to produce the susceptibility map and both model performance are evaluated by Accuracy score, Kappa score and AUC. The final output of landslide susceptibility map will show the susceptibility level in three categories like high, medium and low.

[1]*Graduted, Urban and Rural Planning Discipline, Khulna University, Khulna, Bangladesh, Email: toritchakraborty@gmail.com*
[2]*Graduted, Urban and Rural Planning Discipline, Khulna University, Khulna, Bangladesh, Email: didarul.kuu.14@gmail.com*
[3]*Graduted, Urban and Rural Planning Discipline, Khulna University, Khulna, Bangladesh, Email:sharier3@gmail.com*
Email of Corresponding Author – toritchakraborty@gmail.com

LANDSLIDE SUSCEPTIBILITY MAPPING USING XGBOOST MODEL IN CHITTAGONG DISTRICT, BANGLADESH

Torit Chakraborty, Md Didarul Islam[2] and Md Shaharier alam[3]

Study Area

Chittagong district is located in the south-eastern region of Bangladesh is selected as the study area which has the second largest and only port city inside it.The coordinate of the district is 22.3375°N 91.8389°E.The total area is 5282.92km^2 .As the Chittagong district has high elevation places with the most hilly areas, landslides occur mostly in this district. So that the study area is selected to identify the high, moderate and low risk zone.

Methodology

Data collection , preparation and Causative factors

In this research, eight causative factors were used for landslide susceptibility mapping. These factors were selected from the literature review and considering available data. All the factor data were collected from multiple sources and processed using GIS and RS techniques . Landset-8 Satellite image was used to prepare land use, NDVI, and distance from river map . SRTM DEM 30m data was used to prepare elevation ,slope, flow accumulation map. Soil type and rainfall data are collected from geological survey of Bangladesh and Bangladesh meteorological department. All the factor maps were prepared in GIS environment.

For training and validation of model, 60 landslide locations(Ahmed,2014),(Rahman,2016) were used, which randomly divided into two groups:70% for training and 30% for validation. Later, the training data sets were used to build a spatial relationship between the factors and landslide locations. Then the model used on validation datasets to justify models actual performance in real-world scenario.The factors maps are shown below (Figure-1).

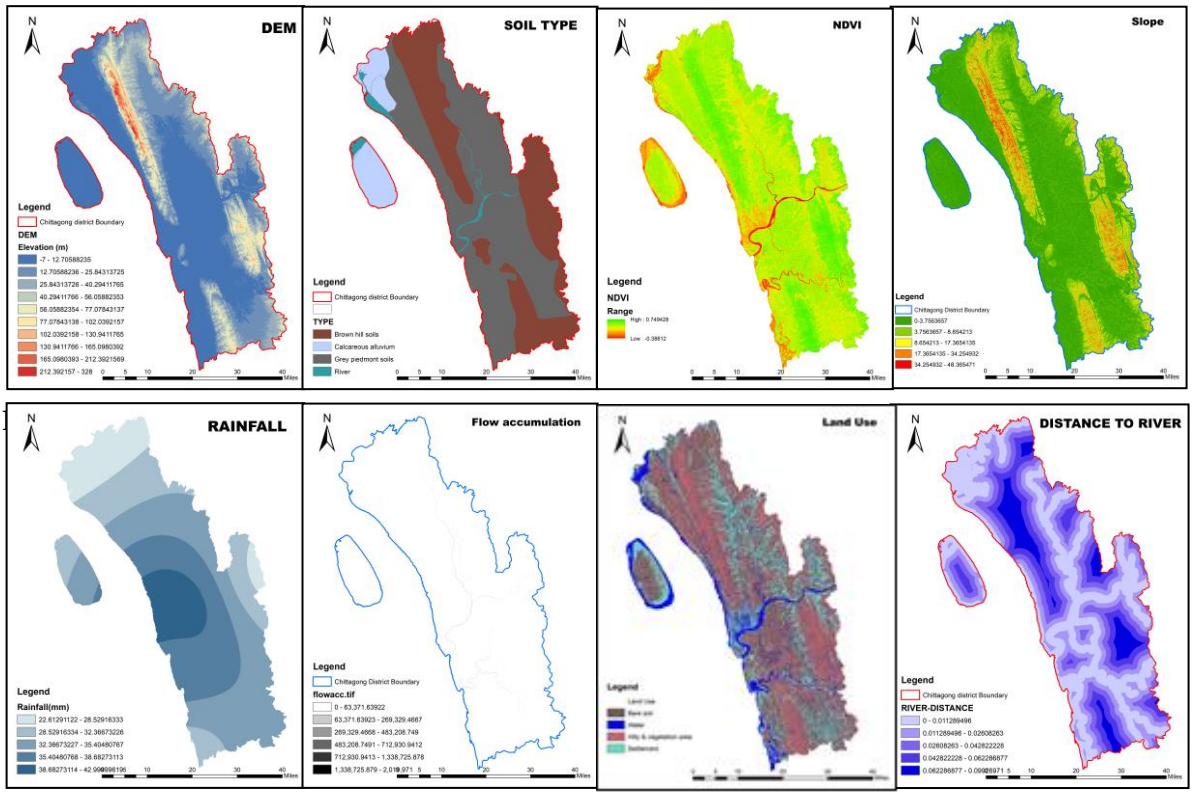

Figure-1. Causative factor maps for Landslide Susceptibility Map

The factors maps show the range and spatial variation of values for the study area. Elevation range is -7 to 328 m where highest elevation found in most of the place. Three types of soil found in the study area where

Grey piedmont soil found most frequently then brown hill soil in the hill area but very little of calcareous alluvium soil. NDVI shows the vegetation index range from -0.38 to 0.74 where negative values indicate water and positive values indicate to different vegetation and built up area. The slope map shows the degree range from 0 to 48 degree where most of the place is below 4 degree but a significant hilly portion has high slope value more than 30 to 48. Flow accumulation map shows the value distribution of water based on surrounding elevation cells which is calculated in ESRI ArcGIS. Rainfall map shows the rainfall distribution of the study area based on the multiple rainfall station data where lowest rainfall rate is 2200 mm and highest 4200 mm. The heavy rainfall pattern found at most of the landslide location. The distance from river map is produced by calculating Euclidean distance of each cells from the river. The land use map shows the land-use distribution in major classes including built-up area, water body, hilly and green area.

Data Analysis

The whole process of data modelling using Xgboost and SVM are performed in the Anaconda platform of python programming language. All the causative factors map and background spatial statistics are made and calculated by ESRI ArcGIS 10.4. The Landsat image layer stacking, NDVI calculation, Land use image classification are done by the Erdas Imagine.
The process of producing landslide susceptibility mapping:

- At the very first stage all the causative factors data are converted to raster format mainly in .TIFF file in ArcGIS.
- After that, all the 60 landslide location data are applied on all the eight causative factor raster map to extract the feature attribute of the corresponding raster cells to prepare training data for the model.
- Then the raster maps are converted into .CSV file (Excel file) to analyze and model the data in Anaconda platform of python programming language.
- Then the training datasets are divided randomly into two groups:70% as training data and 30% as validation data using random split function.
- After training the datasets, both modelling algorithms are applied on validation datasets.
- The validation data are used to check the performance of both model by using Accuracy score, Kappa score and Area under curve score statistics.
- Finally, the model applied to whole dataset to classify the study area into three landslide susceptible categories: low susceptible, medium susceptible and high susceptible.

The work flow of the modeling process is shown in figure-2.

Figure-2. Working process flow of Landslide Susceptibility Mapping

Result and Discussion

Figure-3 shows the relative importance of each factors in constructing model. The graph shows soil type and rainfall is the most influential factors for landslide whereas flow accumulation is less influential. After the training, both models are evaluated by Accuracy score, Kappa score and Area Under Curve(AUC) to assess the performance. The performance of both models are given below:

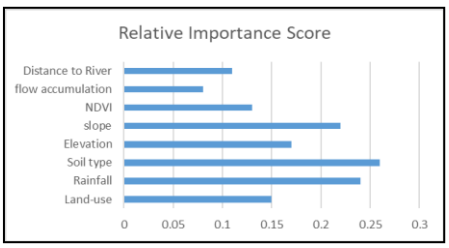

Figure-3: Relative Importance of Each Factors in Landslide Susceptible modeling

Table.1-Validation Performance of two model

Model	Accuracy Score	Kappa	AUC
XGBoost	94.53%	93.63%	88.23%
SVM	91.02%	89.77%	84.38%

Torit Chakraborty, Md Didarul Islam[2] and Md Shaharier alam[3]

According to the table-1, the XGBoost model performs better than SVM model in-terms of validation datasets. The XGBoost model gain 93.63 % kappa score and 94.53 % accuracy score respectively whereas SVM model gain 89.77% kappa score and 91.02% accuracy score. In terms of AUC, XGBoost model also gain higher accuracy than SVM model. So, based on the performance, it can be concluded that the XGBoost model is more suitable than traditional SVM model for mapping landslide susceptibility. Finally, the XGBoost model was applied on the study area to map the landslide susceptible area (Figure-4). After area calculation of each category of landslide susceptibility, we found that 32.47% of the study area is highly susceptible, 19.75% area is moderately susceptible and 47.75% area is low susceptible.

Conclusion

As landslide is a regular phenomenon is the study area which cause casualties and damage of live and wealth, mapping landslide susceptible area can be an effective alternative measure to reduce causalities. But modeling the landslide susceptible area is a complex process which require well understanding of the study area and the factors responsible for landslide. Besides, selection of landslide modeling techniques also an important criterion for effectively mapping the most susceptible landslide area. One of the main objective of this study is to identify the most responsible factors for landslide and selection of suitable modeling techniques. Based on the above analysis, we found that the soil type and rainfall are the most influential factors for landslide. Besides, we also found that the recently developed machine learning algorithm XGBoost perform better than traditional SVM model for mapping the landslide susceptible area.

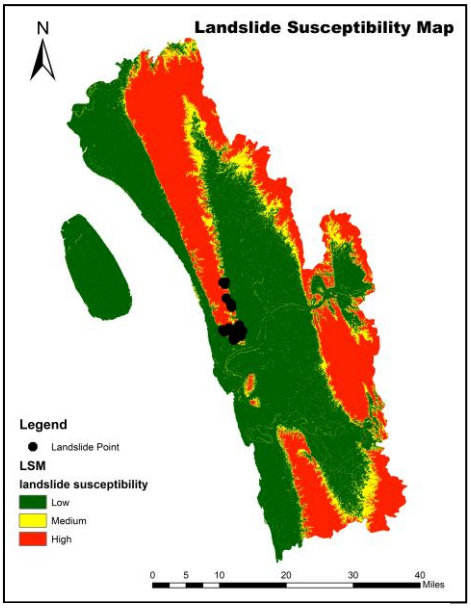

Figure-4. Landslide susceptible map using XGBoost model

References

- Ahmed, B. (2015). Landslide susceptibility modelling applying user-defined weighting and data-driven statistical techniques in Cox's Bazar Municipality, Bangladesh. *Natural Hazards*, *79*(3), 1707-1737. doi: 10.1007/s11069-015-1922-4
- Dong, H., Xu, x., Wang, L., & Pu, F. (2018). Gaofen-3 PolSAR Image Classification via XGBoost and Polarimetric Spatial Information. Sensors, 18(2), 611. doi: 10.3390/s18020611
- Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. IEEE Geoscience And Remote Sensing Letters, 15(4), 607-611. doi: 10.1109/lgrs.2018.2803259
- Huang, Y., & Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. CATENA, 165, 520-529. doi: 10.1016/j.catena.2018.03.003
- Rahman, M.S.; Ahmed, B.; Huq, F.F.; Rahman, S.; Al-Hussaini, T.M. (2016). Landslide Inventory in an Urban Setting in the Context of Chittagong Metropolitan Area, Bangladesh. Paper Id: 36 (pp. 170–178), 3rd International Conference on Advances in Civil Engineering 2016 (ICACE 2016), 21–23 December 2016, Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh.
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., & Pourghasemi, H. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. Computers & Geosciences, 112, 23-37. doi: 10.1016/j.cageo.2017.11.019