8-2021

Landslide Mapping and Susceptibility Assessment of Chittagong Hilly Areas, Bangladesh

Yasin Wahid Rabby
University of Tennessee, Knoxville, yrabby@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Yasin Wahid Rabby entitled "Landslide Mapping and Susceptibility Assessment of Chittagong Hilly Areas, Bangladesh." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Geography.

Dr. Yingkui Li, Major Professor

We have read this dissertation and recommend its acceptance:

Dr. Kelsey Ellis, Dr. Ed Perfect, Dr. Jon M. Harbor, Dr. Haileab Hilafu

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Landslide Mapping and Susceptibility Assessment of Chittagong Hilly Areas, Bangladesh

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Yasin Wahid Rabby

August 2021

Acknowledgment

I want to express my appreciation to my Ph.D. advisor, Dr. Yingkui Li, for supporting me throughout my Ph.D. life. This research would not have been possible without the continuous advice, guidance, and inspiration of Dr. Li. He was always careful about my progress and success in my research. I felt very lucky that I was supervised by Dr. Li, who was excellent at field analysis, GIS, and remote sensing. He was a down-to-earth personality and was always ready to help me so that I can excel in my academic life. He has given me complete freedom to choose this research topic and has allowed me to work on various projects to develop my expertise. He is a very knowledgeable person and always came with an insightful discussion on my research. He has always ensured that my research is going in the right direction. To expedite my research, he trained me in Google Earth-based and landslide mapping and field mapping methods. Without this training, the first chapter of this research would not have been possible to finish. He was very cautious about revising my manuscript and other scholarly writings.

I will forever be thankful to my great committee members, Drs Jon Harbor, Ed Perfect, Kelsey Ellis, and Haileab Hilafu. Their contribution and suggestions have helped me in polishing my dissertation. They were very cooperative and always ready to help in improving the quality of my dissertation. I am thankful to Dr. Haileab and Ed Perfect for their suggestions in statistical analyses. I am grateful to Dr. Kelsey Ellis for her advice on the uncertainty analyses of the landslide inventory. Finally, the advice of Dr. Jon Harbor enabled me to host the data of the landslide inventory on an open-source platform.

I am thankful to the University of Tennessee, Knoxville, since I was awarded the McClure scholarship and Penley Thomas fellowship. This funding helped me to carry out fieldwork in Bangladesh as well as to buy necessary instruments.

I am indebted to my friends' support from the University of Tennessee Knoxville and the University of Dhaka. They are Kyle Landolt, Zach Merril, Joynal Abedin, Mostafizur Rahman, Mostafa Amir Foysal, and Ahmed Saqib Antor. I am very grateful to Jacob Cecile for proofreading and editing my writing. Finally, I am indebted to my mother for her sacrifice, patience, and unconditional love during my Ph.D. life. She did not understand my research but always was interested to know about it. She is an inspiration during my failure and a source of comfort to me, although she was a thousand miles away from me.

Abstract

Landslides are natural phenomena in mountainous areas that cause damage to properties and death to people around the world. In Bangladesh, landslides have caused enormous economic loss and casualty in Chittagong Hilly Areas (CHA). In this dissertation, a landslide inventory of CHA was prepared using Google Earth and field mapping. Google Earth-based mapping helped in recording landslides in inaccessible areas like forests. In contrast, field mapping helped in mapping landslides in accessible areas like areas near road networks. This research also proposed a Mahalanobis distance (MD) based absence-data sampling method to objectively select non-landslide locations for landslide susceptibility mapping. This proposed method was demonstrated in the landslide susceptibility mapping of the three Upazilas (subdistricts) of Rangamati district, Bangladesh, and the generated landslide susceptibility map was compared with the map produced by the slope-based absence data sampling. Fifteen landslide causal factors, including slope aspect, plan curvature, and geology, were used in the random forest model for landslide susceptibility mapping. The areas under the success and prediction rate curves, as well as statistical indices, showed that both absence-data sampling methods provided similar accuracy, but the seed cell area index (SCAI) showed that MD based landslide susceptibility map is more consistent and does not overestimate the landslide susceptibility like the slope-based model. Finally, this dissertation research assessed the impact of three land use/land cover (LULC) scenarios (a. existing (2018); b. proposed LULC (Planned); and c. simulated (2028) LULC) on the landslide susceptibility of the Rangamati municipality using the random forest model. The results showed that high susceptibility zones would increase in both proposed and simulated LULC scenarios, but the increase is comparatively low in the proposed LULC. Although the proposed LULC scenario did not consider landslide susceptibility, the implementation of general LULC planning rules, such as avoiding sleep slopes for road and build-up constructions, helped to mitigate landslide susceptibility.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Research Overview

Landslides refer to the movement of debris, rocks, soil, and earth under the influence of gravity (Cruden and Varnes, 1996). It is a naturally occurring phenomenon in mountainous areas (Roy and Saha, 2019) and accounts for 9% of the natural disasters in the world (Galli et al., 2008; Kanwal et al., 2016). Landslides cause damage to infrastructure, leading to human fatalities and economic losses (Guzzetti et al. 2000; Yilmaz, 2009; Chen et al., 2017; Wang et al., 2017). For example, it caused the death of 8739 people and affected 3.2 million people directly and indirectly from 2004 to 2013 (Ahmed and Dewan, 2017).

Landslides are affected by causal and triggering factors. Causal factors create a suitable condition for landslides, whereas triggering factors initiate the landslides (Guzzetti et al., 2012). The causal factors of landslides include slope, aspect, curvature, geology, and land use/land cover (Ahmed 2015). Landslides can be triggered naturally by snow melting, volcanic activity, groundwater pressure, and prolonged rainfall (Guzzetti et al., 2012; Arora et al., 2014; Chen et al., 2017). Landslides can also be triggered by human activities, such as excavation, deforestation, land-use change, hillslope cutting, construction of roads, and subsequent excessive vibration by traffic and agricultural cultivation (Althuwaynee et al., 2014; Althuwaynee et al. 2016; Chen et al., 2017).

Landslide inventory and susceptibility mapping have been argued as the first two steps towards landslide assessment (Guzzetti et al. 2006; Guzzetti et al. 2009; Guzzetti et al. 2012; Kanwal et al. 2016; Chen et al. 2017). Landslide inventory shows the locations of landslides that occurred in the past and can be used to produce and validate landslide susceptibility maps (Zezere et al., 2017). Landslide causal factors are also critical for landslide susceptibility mapping (Ahmed, 2015; Ahmed et al., 2018). Detailed analysis of landslide causal factors at landslide locations is useful to determine the likelihood of landslides over an area and produce the susceptibility maps (Yilmaz, 2009; Yilmaz, 2010; Sterlacchini et al. 2011).

1.1.1 Landslide Inventory Maps

A landslide inventory map shows the locations and distribution of landslides that have left discernible traces over an area (Guzzetti et al., 2012). It contains different attributes, such as type, extent, location of occurrence, information about the surrounding area, and landslides' damage (Guzzetti et al., 2006; Guzzetti et al., 2012). Depending on the mapping scale, landslides can be

represented as a point or an area. Landslide inventory provides a snapshot of the landslides during a given period but may not show the evolution of landslides in the long term. Landslide inventory documents the extent, type, and causes of landslides, helping prepare and validate the susceptibility models (Guzzetti et al. 2006; Ahmed and Dewan, 2017).

Mapping landslide inventory depends on the scale and mapping purpose (Guzzetti et al., 2012). Medium to large scale (<1:10000) landslide inventories can be derived from the interpretation of high-resolution aerial photographs, satellite imagery, and extensive field mapping (Guzzetti et al. 2002). Small scale (>1:100000) landslide inventories can be documented based on literature, newspaper, journals, technical and scientific reports, governmental reports, and the interview of experts (Glade, 2001).

Traditional methods in landslide inventory mapping are mainly based on field mapping and visual interpretation of aerial images, topographic maps, printed maps, and archives or reports (Alkevli and Ercanoglu, 2011). Automated and semi-automated mapping techniques and interpretation of digital images are also developed based on the analysis of very high-resolution Digital Elevation Model (DEM), interpretation of high or medium optical remote sensing data, and analysis of Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR) data (Guzzetti et al. 2012). All methods have advantages and disadvantages. Field mapping ensures a better assessment, but it is time-consuming, and some remote places are inaccessible (Alkevli and Ercanoglu, 2011). Aerial photographs cover large areas, but their interpretation may be subjective, and the accuracy of the interpretation depends on the experience and skills of the interpreter and the quality of the stereoscope (Alkevli and Ercanoglu, 2011; Guzzetti et al. 2012).

1.1.2. Landslide Susceptibility Mapping

Landslide susceptibility map shows the probability of landslides over an area. It uses previous landslide locations and their relationship with the causal factors to predict the likelihood of future landslides (Ayalew and Yamagishi, 2005). The principle of landslide susceptibility mapping assumes that future landslides will occur in areas where geo-environmental conditions are similar to where landslides previously occurred (Guzzetti et al., 2012).

Landslide susceptibility can be investigated using quantitative and qualitative methods. Quantitative methods determine the relationship between landslides' locations and their associated causal factors (Althuwaynee et al., 2014). These methods are limited by the oversimplification of

causal factors. Quantitative methods can be categorized as deterministic and statistical methods. In a deterministic approach, a safety factor is commonly defined based on a few causal factors to determine the landslide susceptibility of an area (Yilmaz, 2009). It is suitable for small areas due to the challenge of measuring the safety factor over a large area (Ayalew and Yamagishi, 2005). Statistical methods can be either bivariate or multivariate (Vakshoori and Zare, 2016). Bivariate techniques compare landslide locations with each causal factor. In this method, each causal factor is divided into a set of classes using user-defined methods, such as natural break or equal interval. Bivariate methods consider the relationship between landslide locations and divided classes of each causal factor. For example, we can divide slopes into several classes and derive the relationship between landslide occurrence and slope classes. Then, we can repeat the same method for other factors (Althuwaynee et al., 2013). In summary, the bivariate models only assess the relationship between landslide occurrence and one factor at a time, although landslides are controlled by a combination of multiple factors (Ayalew and Yamagishi, 2005). The commonly used bivariate methods include frequency ratio, the weight of evidence, fuzzy logic, evidential belief function, and statistical index (Vakshoori and Zare, 2016; Chen et al., 2017). The multivariate statistical methods determine the relationship between landslide occurrence and multiple causal factors. Examples of multivariate methods include logistic regression, adaptive regression spline, general additive models, and simple decision trees. These methods can outperform the bivariate and multivariate methods (Yilmaz, 2010) but usually lack the power of interpretability (Althuwaynee et al., 2014).

Qualitative methods depend on expert knowledge and judgment. Examples of qualitative methods include the Analytical Hierarchy Process (AHP) and Weighted Linear Combination (Yilmaz, 2009). These methods are mainly based on the weights of causal factors that are subjectively assigned based on expert knowledge and then combine the weighted value of each factor to produce the susceptibility map (Kanwal et al., 2016).

The selection of methods for landslide susceptibility mapping depends on the scale, cost, and timeline of the analysis (Yilmaz, 2009). For instance, deep learning techniques like Artificial Neural Networks (ANN) show high predictive capability but require time and high computational power (Akgun et al., 2012). Bivariate analysis requires an inventory that covers the whole area because its produced landslide susceptibility maps follow the known landslide locations. However,

it is impossible to map all landslides in a complicated terrain; thus, the produced susceptibility maps can be biased towards the known landslide locations (Schicker and Moon, 2012; Petschko et al., 2014). Multivariate models like logistic regression have generalization capacity, and results are easily interpretable (Akgun et al., 2012). It is up to the researchers to compare different models and determine which one is the best for a specific area (Vakshoori and Zare, 2016).

Appropriate model selection for regional and national susceptibility mapping requires prudent judgments. These maps are created for regional planning and land use management (Sabatakakis et al., 2012; Schicker and Moon, 2012). It is necessary to select a proper sampling strategy, factors, and methods. Bivariate models do not require non-landslide locations, while multivariate and machine learning methods require the sampling of both landslide and non-landslide locations. If the selection of non-landslide locations is not representative, the susceptibility maps would be biased towards specific geomorphic or topographic units (Chen et al., 2019).

In recent years, the use of integrated or hybrid models has increased to reduce the variance and increase the prediction capability (Althuwaynee et al., 2014; Li et al., 2019). Hybrid models can integrate bivariate models with multivariate, machine learning, and qualitative models (Althuwaynee et al., 2016). Althuwaynee et al. (2014) integrated evidential belief function (EBF), a bivariate model with analytical hierarchy process (AHP) and logistic regression for Pohen and Gyeongju cities of South Korea. This integration reduced subjectivity and increased prediction capability to 80 - 82.3%. However, their study area was relatively small; thus, the question remains whether the integration can produce better predictions for large areas. Xu et al. (2019) integrated the index of entropy with logistic regression and support vector machine for Shaanxi Province of China. Their results indicated that the integration with the logistic regression provided a better prediction than the integration with support vector machines. Some studies suggested that integrating bivariate and multivariate models produces better results than the integration of bivariate and machine learning models (Althuwaynee et al., 2014). Chen et al. (2018) integrated three bivariate models of the index of entropy, certainty factor, and statistical index with a machine learning method of random forest from Shaanxi province of China. This study suggested that the integration of certainty factor with random forest shows better prediction capability. Althuwaynee et al. (2016) integrated the chi-squared automatic interaction detection with AHP and suggested this integrated approach outperforms the AHP method. Rossi et al. (2010) introduced an optimal

landslide susceptibility model by combining two or three models. They did not integrate the models during the building stage. Instead, they produced the susceptibility maps for each model and then integrated them as the optimal model using a regression-based approach. They compared the optimal model results with the ones produced using linear discriminant analysis, quadratic discriminant analysis, and logistic regression and indicated that the optimal model produced the best prediction among these models.

1.1.2.1 Sampling non-landslide locations

Most statistical models and machine learning methods require both landslide and non-landslide locations for landslide susceptibility mapping. Landslide locations are derived from the landslide inventory, while the determination of non-landslide locations requires certain sampling strategies. Random sampling is the most common approach to choose a non-landslide location. The assumption is that all locations other than the landslides can be considered non-landslide locations (Tsangaratos and Benardos, 2014; Regmi et al., 2014). Some studies used data exploratory analysis to select a safe zone (where the chance of landslides is minimum), and non-landslide locations are selected randomly from this area (Althuwaynee et al., 2014). Data exploratory analysis often brings bias to the susceptibility maps. For instance, if a safe zone is selected based on slope, the results will be biased to the slope (Hong et al., 2019). The proportion of landslide and non-landslide locations is an important factor for multivariate and machine learning methods, and it can be 1:10, 1:5, 1:2, and 1:1 (Othman et al. 2018). Heckmann et al. (2014) opined that the 1:1 method gives the best prediction.

1.1.2.2 Selection of Causal Factors

The quality and plausibility of landslide susceptibility maps depend on the quality of landslide inventory and causal factors (Budimir et al., 2015). The selection of causal factors depends on the availability of data, timeline, cost of the project, and size of the study area (Remondo et al. 2003). DEM is essential data for the determination of causal factors. Different topographic factors, such as slope, aspect, topographic wetness index (TWI), and stream power index (SPI), are generated from DEM using GIS (Marchesini et al. 2014). Free satellite images like the Landsat series are used to prepare land use/land cover and normalized difference vegetation index (NDVI) maps (Ahmed, 2015). Several studies have classified these factors into different categories (Budimir et al., 2015). Kanwal et al. (2016) classified causal factors into four groups: a) human-induced parameters, including land use/land cover and road density; b) topographic parameters, including

slope, aspect, and curvature; c) hydrological parameters, such as river network, SPI, and TWI; and d) geology, including lithology and fault lines. Reichenbach et al. (2018) divided causal factors into five clusters: a) morphological; b) geological; c) land cover; d) hydrological and e) other variables. It is recommended to take at least one factor from each of the groups for landslide susceptibility mapping (Budimir et al., 2015).

Commonly used causal factors include slope, aspects, curvature, distance to the road network, river network, fault lines, land use/land cover, TWI, and SPI (Budimir et al. 2015). Reichenbach et al. (2018) opined that distance to linear features like road networks often brings biases to the model. The landslide susceptibility maps follow the pattern of mapped landslides (Guzzetti et al., 2012).

In bivariate models (other than the weight of evidence), causal factors cannot be selected based on their significance. All the factors are included in the model, and multicollinearity is not considered, leading to biases and poor prediction capability. This is a problem for regional and national scale landslide susceptibility mapping (Regmi et al., 2014). Multicollinearity is usually considered in multivariate and machine learning methods, producing more plausible results.

1.1.2.3 Model Evaluation

Model fit and prediction performances are used to evaluate the susceptibility maps (Rossi et al. 2010). During model formulation, landslide locations are divided into two sets: training and validation sets (Yilmaz, 2009). The training set is used to test how well the model describes the known landslide locations. Validation sets are used to test how well the model can predict the unknown landslides (Frattini et al. 2010). The partitioning of the dataset can be based on different ratios. Most studies use either 80:20 or 70:30 ratios (Sabokbar et al., 2014). The receiver operating characteristics (ROC) curves are used to show success and prediction performance. For ROC curves, the larger the area under the curves (AUC), the better the model performance (Vakshoori and Zare, 2016; Shirzadi et al., 2017; Zhu et al. 2019). Relative density index, frequency measures, and confusion matrices are also used for model evaluation (Guzzetti et al. 2006). Different model evaluation methods have their specific advantages and disadvantages. It is recommended to use multiple evaluation metric to assess the model performance (Rossi et al. 2010).

1.2. Objectives and significance of this study

This dissertation research focuses on mapping landslides, proposing an objective absence-data sampling method for landslide susceptibility mapping, and evaluating the impact of land use/land cover change on landslide susceptibility. The study area is the Chittagong Hilly Areas of Bangladesh. The detailed objectives are:

1. To map all known landslide locations of CHA using field mapping and Google Earth mapping.
2. To evaluate the Mahalanobis distance (MD)-based absence-data sampling or non-landslide location selection for landslide susceptibility mapping.
3. To evaluate the effects of different land use and land cover scenarios on landslide susceptibility.

Landslides are the third deadliest disaster in the world (Ahmed, 2015). In recent decades, human activities have expanded to mountainous areas due to population growth and tourism development. This reduced the slope stability, contributing to an increase in landslides (Guzzetti et al., 2012). Landslide inventory and susceptibility mapping are essential for urban and regional planning to take precautionary measures in the landslide-prone areas.

Landslides are common hazards in the CHA, but CHA does not have a landslide inventory except for the two urban areas of the Chittagong Metropolitan Area (CMA) and Cox's Bazar municipality. This study provided the first landslide database of CHA. Field mapping is the most widely used method for landslide inventory mapping, but this method can only be applied to accessible areas (Fell et al. 2008). To ensure both the accessible and inaccessible areas are covered for landslide inventory mapping, this study integrates field mapping with the Goggle Earth image interpretation to map landslides in CHA. This inventory can be used for landslide susceptibility mapping for the entire CHA.

Landslide susceptibility mapping requires both presence (landslides) and absence (non-landslide locations) data (Zhu et al., 2019); however, the selection of absence-data is usually subjective. This research proposed an objective MD-based absence-data sampling based on a theoretical Chi-square distribution of MD values and a specific confidence level. This method was then compared with a traditional slope-based absence-data sampling method to evaluate the model

performance, accuracy, and consistency in the landslide susceptibility mapping of three Upazilas of Rangamati district, Bangladesh.

Most landslide causal factors, such as slope, aspect, and geology, are relatively stable and static. Anthropogenic factors like land use/land cover can frequently change in areas like CHA where people live in the foothills and change the slope structure for different development activities. The dynamics of land use/land cover change may affect the susceptibility of landslides. This study assessed the contribution of land use/land cover change on landslide susceptibility. This work would provide a useful guidance for land use planning in landslide-prone areas**.**

1.3. Dissertation organization

This dissertation is organized in a manuscript format that includes three manuscripts targeted for different journals.

Chapter 2 focuses on mapping landslides in the CHA, Bangladesh. A total of 730 landslides were mapped based on the integration of field mapping and Google Earth mapping. These landslides occurred between 2001 to 2017. Google Earth mapping helped cover inaccessible areas like the forests, and field mapping helped cover accessible areas such as the urban areas to map the landslides in the study area.

The proposed MD-based absence-data sampling method for landslide susceptibility mapping was described in Chapter 3 with a comparison of a commonly used slope-based absence-data sampling. Three Upazilas (subdistrict) of Rangamati district, Bangladesh, were used as the test site. Fifteen landslide causal factors, including slope aspect, elevation, plan curvature profile curvature, distance from the drainage network, and rainfall and 261 landslide locations were used in calculating the MD and later compared with the Chi-square distribution to determine a threshold above which safe zone for absence-data sampling can be defined. The random forest model was used for landslide susceptibility mapping, for accuracy assessment and consistency analysis, and to compare the effects of MD and slope-based absence-data sampling on landslide susceptibility mapping, success and prediction rates, statistical indices, including the Kappa values and seed cell area index were used.

Chapter 4 presents the work to evaluate the impact of land use/landcover (LULC) on landslide susceptibility maps in the Rangamati municipality of Rangamati district, Bangladesh, based on

three LULC scenarios: the existing LULC (2018); a proposed LULC (planned); and a simulated (2028) LULC. The random forest model was used in landslide susceptibility mapping, and success and prediction rates were used for accuracy assessment. The overall correlation was used in assessing the correlation among the three landslide susceptibility maps. Spatial and areal comparisons were used to determine whether the planned and simulated LULC increased the study area's landslide susceptibility.

Chapter 5 summarizes the findings of the landslide inventory mapping in CHA, MD-based absence-data sampling method, and the impact of LULC on the landslide susceptibility map. It also discusses the potential future work regarding the landslide inventory and susceptibility mapping.

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), pp.1707-1737.

Ahmed, B. and Dewan, A., 2017. Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in Chittagong City Corporation, Bangladesh. *Remote Sensing*, *9*(4), p.304.

Ahmed, B., Rahman, M., Islam, R., Sammonds, P., Zhou, C., Uddin, K. and Al-Hussaini, T.M., 2018. Developing a dynamic web-gis based landslide early warning system for the chittagong metropolitan area, Bangladesh. *ISPRS International Journal of Geo-Information*, *7*(12), p.485.

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C. and Pradhan, B., 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Computers & Geosciences*, *38*(1), pp.23-34.

Alkevli, T. and Ercanoglu, M., 2011. Assessment of ASTER satellite images in landslide inventory mapping: Yenice-Gökçebey (Western Black Sea region, Turkey). *Bulletin of Engineering Geology and the Environment*, *70*(4), pp.607-617.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides*, *11*(6), pp.1063-1078.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena*, *114*, pp.21-36.

Althuwaynee, O.F., Pradhan, B. and Lee, S., 2016. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. *International Journal of Remote Sensing*, *37*(5), pp.1190-1209.

Arora, M.K., Das Gupta, A.S. and Gupta, R.P., 2004. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. International Journal of Remote Sensing, 25(3), pp.559-572.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, *65*(1-2), pp.15-31.

Budimir, M.E.A., Atkinson, P.M. and Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. *Landslides*, *12*(3), pp.419-436.

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I. and Dick, O.B., 2012. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. *Geomorphology*, *171*, pp.12-29.

Chen, W., Zhang, S., Li, R. and Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Science of the total environment*, *644*, pp.1006-1018.

Chen, W., Shahabi, H., Shirzadi, A., Hong, H., Akgun, A., Tian, Y., Liu, J., Zhu, A.X. and Li, S., 2019. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment, 78(6), pp.4397-4419.

Cruden, D.M. and Varnes, D.J., 1996. Landslides investigation and mitigation. Landslide types and processes. *Special report*, *247*.

Frattini, P., Crosta, G. and Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. *Engineering geology*, *111*(1-4), pp.62-72.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology*, *94*(3-4), pp.268-289.

Glade, T., 2001. Landslide hazard assessment and historical landslide data—an inseparable couple?. In *The use of historical data in natural hazard assessments* (pp. 153-168). Springer, Dordrecht.

Guzzetti, F., 2002, October. Landslide hazard assessment and risk evaluation: Limits and prospectives. In *Proceedings of the 4th EGS Plinius Conference, Mallorca, Spain* (pp. 2-4).

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M., 2006. Estimating the quality of landslide susceptibility models. *Geomorphology*, *81*(1-2), pp.166-184.

Guzzetti, F., Ardizzone, F., Cardinali, M., Rossi, M. and Valigi, D., 2009. Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth and Planetary Science Letters*, *279*(3-4), pp.222-229.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), pp.42-66.

Heckmann, T., Gegg, K., Gegg, A. and Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Natural Hazards and Earth System Sciences, 14(2), p.259.

Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W. and Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, pp.399-413.

Kanwal, S., Atif, S. and Shafiq, M., 2017. GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomatics, Natural Hazards and Risk*, *8*(2), pp.348-366.

Li, C., Yan, J., Wu, J., Lei, G., Wang, L. and Zhang, Y., 2019. Determination of the embedded length of stabilizing piles in colluvial landslides with upper hard and lower weak bedrock based on the deformation control principle. *Bulletin of engineering geology and the environment*, *78*(2), pp.1189-1208.

Marchesini, I., Ardizzone, F., Alvioli, M., Rossi, M. and Guzzetti, F., 2014. Non-susceptible landslide areas in Italy and in the Mediterranean region. *Natural Hazards and Earth System Sciences*, *14*(8), pp.2215-2231.

Othman, A.A., Gloaguen, R., Andreani, L. and Rahnama, M., 2018. Improving landslide susceptibility mapping using morphometric features in the Mawat area, Kurdistan Region, NE Iraq: Comparison of different statistical models. *Geomorphology*, *319*, pp.147-160.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T., 2014. Assessing the quality of landslide susceptibility maps–case study Lower Austria. *Natural Hazards and Earth System Sciences*, *14*(1), pp.95-118.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, pp.60-91.

Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T. and Akgun, A., 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences*, *7*(2), pp.725-742.

Remondo, J., González, A., De Terán, J.R.D., Cendrero, A., Fabbri, A. and Chung, C.J.F., 2003. Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain. *Natural Hazards*, *30*(3), pp.437-449.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C. and Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology*, *114*(3), pp.129-142.

Roy, J. and Saha, S., 2019. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenvironmental Disasters*, *6*(1), pp.1-18.

Sabokbar, H.F., Roodposhti, M.S. and Tazik, E., 2014. Landslide susceptibility mapping using geographically weighted principal component analysis. *Geomorphology*, *226*, pp.15-24.

Sabatakakis, N., Koukis, G., Vassiliades, E. and Lainas, S., 2013. Landslide susceptibility zonation in Greece. *Natural hazards*, *65*(1), pp.523-543.

Schicker, R. and Moon, V., 2012. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology*, *161*, pp.40-57.

Shirzadi, A., Shahabi, H., Chapi, K., Bui, D.T., Pham, B.T., Shahedi, K. and Ahmad, B.B., 2017. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. Catena, 157, pp.213-226.

Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M. and Sorichetta, A., 2011. Spatial agreement of predicted patterns in landslide susceptibility maps. *Geomorphology*, *125*(1), pp.51-61.

Vakhshoori, V. and Zare, M., 2016. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomatics, Natural Hazards and Risk*, *7*(5), pp.1731-1752.

Wang, Q., Wang, Y., Niu, R. and Peng, L., 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*, *9*(9), p.938.

Xu, L., Coop, M.R., Zhang, M. and Wang, G., 2018. The mechanics of a saturated silty loess and implications for landslides. *Engineering Geology*, *236*, pp.29-42.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences*, *35*(6), pp.1125-1138.

Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environmental Earth Sciences*, *61*(4), pp.821-836.

Zêzere, J.L., Pereira, S., Melo, R., Oliveira, S.C. and Garcia, R.A., 2017. Mapping landslide susceptibility using data-driven methods. *Science of the total environment*, *589*, pp.250-267.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T. and Hong, H., 2019. A similarity-based approach to sampling absence-data for landslide susceptibility mapping using data-driven methods. *Catena*, *183*, p.104188.

Chapter 2

An Integrated Approach to Map Landslides in Chittagong Hilly Areas, Bangladesh, using Google Earth and Field Mapping

This chapter is a manuscript and published in Landslides journal. According to the author's guidelines, it is mandatory to have a separate figure and table files. The format of this chapter follows the requirements of this journal. The use of "we" in this chapter refers to co-author, Dr. Yingkui Li, and me. As the first author, I did the analysis and wrote the manuscript.

Abstract

This paper presents a landslide inventory map for the Chittagong Hilly Areas of Bangladesh based on Google Earth and field mapping. We developed a set of criteria to identify landslides in Google Earth and introduced a method to assess the accuracy of mapped landslides in Google Earth, which is suitable for the landslides that are mapped as points rather than polygons in the field. In total, 230 landslides (mainly occurred in 2001-2016) were mapped in Google Earth. Field mapping identified 548 landslides that occurred mainly during Summer 2017. The total inventory includes 730 landslides for Chittagong Hilly Areas area from 2001 to 2017. The accuracy assessment suggests that the accuracy of mapped landslides using Google Earth varies from 69-88%. Field work helps to map landslides in urban areas, near to road networks, human settlements, and accessible areas, whereas Google Earth helps to map landslides in inaccessible areas. The combination of these two approaches provides a means to prepare the landslide inventory for an entire area.

2.1. Introduction

Landslides are a common earth surface process in mountainous areas and play an important role in landscape evolution (Galli et al. 2008; Netra et al. 2014). They represent 9% of the natural disasters in the world (Guzzetti et al. 2000), causing damage to infrastructure and loss of lives (Guzzetti et al. 2000; Yilmaz 2009; Netra et al. 2010; Myronidis et al. 2016; Wang and Li 2017; Chen et al. 2017). Landslides can be triggered naturally by rapid snow melting, volcanic activity, groundwater pressure, and prolonged rainfall (Guzzetti et al. 2012; Pandey 2015; Peruccacci et al. 2017; Chen et al. 2017). They can also be triggered by human activities, such as excavation, deforestation, land use change, hill cutting, and road construction agricultural cultivation (Chen et al. 2017).

Landslide inventory mapping is an important step for landslide susceptibility, hazard, and risk assessment (Guzzetti et al. 2012; Kanwal et al. 2016). Landslide can be mapped as a point or a polygon depending on the scale (Guzzetti et al. 2012). Landslide inventory includes archival and geomorphological inventories (Alkevli and Ercanoglu 2011). An archive inventory shows the extent, type and location of landslides. Geomorphological inventories include historical and seasonal or multi-temporal inventories. A historical inventory shows cumulative landslide events over hundred and thousand years. A seasonal inventory shows single or multiple landslide events during a single season or few seasons (Galli et al. 2008).

Various techniques have been used for landslide inventory mapping (Guzzetti et al. 2012). Traditional methods include the interpretation of aerial photographs, satellite imagery and field mapping. These methods are commonly used to generate landslide inventory maps for a large area (Alkevli and Ercanoglu 2011). Data obtained from the literature, newspaper, journals, technical reports, governmental archives, and the interviewing of experts were also used to prepare landslide

inventories for small areas (Glade 1998). In recent years, landslides have also been mapped using high resolution Digital Elevation Model (DEM), Light Detection and Ranging (LiDAR), and Synthetic Aperture Radar (SAR) data (Guzzetti et al. 2012).

Bangladesh is a primarily low-lying floodplain country in South Asia. Mountainous terrain covers only 18% of the land on the north, northeast and southeast. Landslides are common in the hilly regions, especially the Chittagong Hilly Areas (CHA) (Fig. 2.1) in south-eastern Bangladesh (Banglapedia 2015). Most landslides occur during the monsoon season in the CHA due to extreme rainfall events (>40 mm/day) within a short period (2-7 days) (Khan et al. 2012). High cloud cover during this season prevents the identification of landslides from high (0.5-5m) and medium (15-30m) resolution multi-spectral images, such as Landsat imagery. High-resolution aerial photographs and imagery are either not available or not free in this area. In addition, vegetation regrows quickly after a landslide event in sub-tropical areas like CHA and it is challenging to identify the landslide in satellite images or aerial photographs after a few months of landslides (Samodra et al. 2015).

Most landslide inventory projects have focused on the major cities of CHA (Ahmed 2015 and CDMP-II 2012). For example, Ahmed and Dewan (2017) and Ahmed (2015) compiled landslide inventories for Chittagong Metropolitan Area (CMA) and Cox's Bazar municipality and developed different techniques in landslide susceptibility mapping. In contrast, few studies have been conducted outside of these two cities. We used the visual interpretation of multi-temporal imagery in Google Earth and extensive fieldwork to map old and recent landslides in CHA. The inventories identified using these two methods are combined to produce a landslide inventory map. High-resolution multi-temporal Google Earth imagery allows for identifying landslides in remote areas where field mapping is not possible. Several studies have used Google Earth for landslide mapping

(Sato and Harp 2009; Fisher et al. 2012; Vakhshoori and Zare 2016), but no criteria are available for identifying landslides in Google Earth. In this study, we developed six criteria for detecting landslides in Google Earth. In addition, previous studies associated with Google Earth-based landslide mapping have not presented any accuracy assessment (Sato and Harp 2009; Fisher et al. 2012; Vakhshoori and Zare 2016). We also introduced an accuracy assessment method for Google Earth-based landslide mapping.

2.2. Study Area

The Chittagong Hilly Area (Fig. 2.1) (20,957 km^2) is in the southeast Bangladesh (20.46°–23.40° N and 91.27°–92.18° E) and includes five districts: Bandarban, Rangamati, Khagrachari, Chittagong and Cox's Bazar. CHA has tropical monsoon climate with annual rainfall ranging from 2540 mm in north and east to 2794 to 3777 mm in south and west. This area has three distinct seasons: the Dry and Cool Season from November to March; the Hot or Pre-monsoon season from April to May, and the Monsoon or Rainy Season from June to October (Rashid 1978; Banglapedia 2015). About 80% of the landslides occurs between May to September when rainfall is >200 mm per month in this area (Khan et al. 2012).

The hilly area can be divided into the low hill ranges (<300 m) and the high hill ranges (>300 m) (Banglapedia 2015). The low hill ranges are under Dupi Tila and Dihing formation whereas the high hill ranges under Surma and Tipam formation (Fig. 2.2) (Brammer 2012). Most of the areas in west have slope <5° and the areas in the east have slope>30° (Fig. 2.2).

2.3. Data Source

We used Google Earth imagery and an existing landslide database to generate the landslide inventory map. Google Earth contains available Landsat imagery (15m–30m pan-sharpened), orthophotos (0.5–2m), high resolution commercial datasets (SPOT, FORMOSAT-2: 0.5–8m;

World View-1 and World View-2: 0.5–2.5m) (Fisher et al. 2012; Crosby 2012). These datasets provide access to sub-meter resolution images for visual interpretation of landslides (Fisher et al. 2012). Google Earth also provides historical imagery to explore the spatio-temporal landslide changes. The users can also delineate features and save them to KML files in Google Earth (Bailey et al. 2012). Google Earth has been used to delineate landslides and assess their extents and characteristics (Sato and Harp, 2009).

Bangladesh does not have an official database for landslides. Department of Disaster Management of People's Republic of Bangladesh records landslides without detailed locations. Most recorded landslides have the locational information only to the low-level administrative division of Bangladesh, such as name of the village. This record is also not updated regularly and not available online. Comprehensive Disaster Management Programme of the Ministry of Disaster and Relief of Bangladesh provides the detailed landslide inventory for Cox's Bazar and Teknaf municipality areas (CDMP-II, 2012). Rahman et al. (2016) and Ahmed et al. (2014) provided landslide inventory for the Chittagong Metropolitan Area (CMA). These inventories provide GPS coordinates, extent, fatalities, and estimated loss of landslides. Newspaper reports on landslides can be another data source as they give the description of where, when, and why landslides occurred, how many people died, and estimated economic loss. However, these reports lack detailed location and dimension of landslides.

2.4. Method

The methodology includes four steps: 1) visual interpretation of Google Earth imagery; 2) field data collection and mapping; 3) field validation and accuracy assessment; and 4) final map production.

2.4.1. Visual Interpretation of Google Earth Imagery

Due to the availability of Google Earth imagery, landslides were mapped from January 2001 to March 2017. The whole region was divided into 4911 rectangles (3.3 km long and 1.3 km wide) to keep track of mapping and prevent visual interpretation of an image more than once (Fig 3.3.f). These rectangles were created using the Fishnet tool in ArcGIS and then converted to a KML file. We started the mapping from the upper-left rectangle (near the Feni river where the Chittagong District starts) and checked the images from left to right in each rectangle. The landslides were identified in Google Earth based on six criteria: change of vegetation in historical images (vegetation was absent in one image but present in previous images), morphological change in historical images (change detection by comparing two historical images), change of texture and color in historical images, the slope and elevation of suspected areas for landslides, and the presence of debris at the toe of suspected areas.

The historical images were examined to detect changes in vegetation and morphology (Fig 3.3.a and Fig 3.3.b). Landslide can remove or destroy the vegetation of an area to expose bare land. However, open field and harvested paddy field can also appear as bare land in Google Earth. The slope and elevation were measured to separate these different possibilities. The change in slope and elevation from the top to the bottom of a suspected landslide area or the bedrock scarp indicates that landslide process has removed bedrock and vegetation; landslide usually does not occur in a gentle slope (Duric et al. 2017). The Add Path tool in Google Earth was used to check the slope of the bare land. This tool generates the topographic profile along the path (line or polygon) (Bailey et al. 2012). We draw the central line of each bare land (Fig 3.3.c) to examine the slope and elevation change along the profile. When landslide occurs, bedrocks and soils are generally deposited at the toe. This deposit was considered as an indication of landslides. The changes in the

texture and color and the presence of mottling in the image can be considered as the presence of landslide (Guzzetti et al. 2012). We went through Google Earth images using a constant eye altitude of 300 m to check all these changes and identify landslides. Zoom in and out tools were used when the eye altitude was not enough to detect these changes.

In our study, the vegetation change and the presence of bare land were the first two indicators of landslides and then the slope and elevation were measured. After that, the morphological change and the presence of mottling and debris at the toe were checked. The presence of mottling and debris depends on the quality (resolution) of the image, and we did not find them in all mapped landslides because temporal high-resolution images are not available in all areas. Thus, our primary criteria for the landslide identification are the presence of bare land, change in vegetation and morphology, and the measurement of slope and elevation. The presence of debris is optional and increases the mapping confidence when available. We also determined the type of landslide according to Cruden and Varnes (1992) and draw polygon (Fig. 3.3.d) around the scarp and run out (if identifiable). The identification of the landslide type depends on the quality of image and extent of landslides. It was relatively easy to determine the type for a large landslide.

Jhum (Traditional Shifting Cultivation) is a common practice of plantation in CHA. It is a type of rotational farming: one slope of the land is cleared by controlled fire for cultivation and then farmers left the slope to regenerate after few years (Masum 2011). Rotational cultivation is the principal driving force for vegetation removal in hilly forest areas of tropical Asia (Fox et al. 2000). In our method, removal of vegetation is considered as one of the primary indicators of landslides. Thus, in CHA where jhum cultivation is practice, there is a high chance that these areas can be misinterpreted as landslides because these areas become barren land (Fig. 3.3.e) after the

harvesting of crops and remain fellow for next season to regrow vegetation. The availability of historical images in Google Earth helps differentiate areas under jhum cultivation from landslides. We explored the pre- and post-images of the bare land to check the presence of jhum crops in that area. In addition, farmers usually select a rectangular or square slope area for slash-burning and crop cultivation. After harvesting the crops, the area becomes a barren land with the rectangular or square shape. Landslide is a natural process, and its boundary (scarp or run out) is usually irregular.

2.4.2. Field Data Collection and Mapping

Landslide records from local newspaper and existing literature, including published and unpublished articles, thesis, and reports, government documents and archives, and available inventory maps, were used for the field mapping (Table. 2.1). Experts, officials of Disaster Management Department of People's Republic of Bangladesh, city planners, and local political leaders were interviewed to figure out which areas are vulnerable to landslides, why landslides occur, and whether there is any change of the pattern of landslides in the area. We collected newspaper reports on landslides from 1980 to July 2017 at the library of the University of Dhaka. The data collection was mainly based on three Bengali newspapers (The Daily Ittefaq, The Daily Inqilab, and The Daily Prothom Alo) and two English newspapers (The Daily Observer and The Daily Star). We hired four data collectors because digital copies of these newspapers were not available. The data collectors collected the date, time, and locations of landslides, number of death and injured, damage of infrastructures, types and causes of landslides, and so on. Some reports provided the name of the vulnerable areas and the areas where people are living on the foot of excavated hills. These reports helped identify target areas for field investigation and mapping. Local offices of Roads and Highways Department of the People's Republic of Bangladesh

provided the locations of landslides that occurred along the roads during June 2017 under their jurisdictions. Most landslides we collected occurred near roads and human settlements both in rural and urban areas.

We adopted participatory field mapping proposed by Samodra et al. (2015) and used the collected landslides from newspapers and existing literature for field checking and mapping. Most collected data provide the general areas where landslides occurred or are vulnerable to landslides without detailed locations (latitudes and longitudes). We asked local people, political leaders, governmental officials, and aid agencies to help find these locations. The field mapping was carried out from July to August 2017. A GPS receiver (Gramin Trex 20x) with an accuracy of 3-10 m was used to collect the latitude and longitude information of each landslide (Fig. 2.4.c, 2.4.d, and 2.4.f). Chain and tape were used to measure the length and width of the landslide. In some cases, the GPS coordinates were measured 3-10 m away from the landslides because of dangerous field conditions as numerous landslides occurred in June 2017 and were occurring during the field mapping. We measured the distance between the GPS location and the landslide using chain or tape. We checked all collected locations in Google Earth to verify whether they are on the right locations. We did not measure the extent of the landslide in the field due to the lack of topographic maps in this area. Instead, we measured the length and width of each landslide.

A form was used to record time and date (if available), landslide characteristics (if recognizable), land use and land cover type of the area and visually identifiable causes and categorical damage assessment. We first visited each targeted area and then checked with the local people (Fig. 2.4.a and 2.4.b) on whether landslides occurred or not in the area. In some cases, the database from the Department of Disaster showed that landslides occurred in completely flat lands

and local people also failed to remember any landslide events there, indicating that there are errors in the government database. With the help of local guides, we found that some landslides occurred inside the compound of houses. We found all landslides reported by newspapers, indicating that newspaper reports are reliable. Landslides occurred in June 2017 were easily identifiable in the field. All recent landslides occurred within the landslide prone areas identified from newspaper reports. Most landslides we mapped in the field are new landslides due to the numerous numbers of landslides occurred recently. Motorbikes and three wheelers were used to make sure that the survey was conducted as quickly as possible. In average, we mapped about 25 landslides per day. We took photographs of each landslide and its surrounding area to help verify the landslide characteristics that we identified during the field investigation.

2.4.3. Validation and Accuracy Assessment

Several methods are available for the validation and accuracy assessment of the landslide mapping. Carrara et al. (1992) introduced a method based on polygon overlay for the landslide validation and accuracy assessment. This method, however, does not consider the uncertainty, errors, and subjectivities of mapped landslide boundaries. Galli et al. (2008) suggested to use a 100 m buffer around landslide polygons as a threshold to account for the uncertainties and errors in landslide mapping. It treats the landslides (polygons) mapped from satellite imagery and the landslides mapped in the field the same if they are within 100 m. We adopted this buffer distance in our study. However, we mainly recorded the landslides as points in the field, whereas delineated landslides as polygons in Google Earth. It is also not possible to check all Google Earth-identified landslides in the field. We chose three sites (Fig 2.5) for the validation and accuracy assessment. We conducted the field mapping in a test site at Bandarban and compared with landslides that we identified in Google Earth. The next site was the Chittagong Metropolitan Area (CMA). We did not conduct field mapping in the CMA because Rahman et al. (2016) provided 57 landslide

locations. The third site was Cox's Bazar municipality where CDMP-II (2012) provided 77 landslide locations. All landslide points provided by these two reports were not used in our study because some of the landslides in in these reports occurred in 1990 and the oldest landslide that we detected in Google Earth was in 2003. We used the proximity of the landslides from two inventories (landslide points in field mapping and landslide polygons in Google Earth) to assess the accuracy. Specifically, if a landslide mapped in Google Earth is <100 m to the landslide points in the field, we treat them as the matched landslides. The Near tool in ArcGIS was used to determine the nearest distance between the Google Earth-mapped landslides and their closest landslide points recorded in the field. Based on the threshold distance of 100 m, the overall accuracy can be defined as:

$$X = \frac{a}{b}$$

where, X is the overall accuracy, a is the number of landslides mapped in Google Earth that are within 100 m distance from landslide points recorded in the field, and b is the total number of landslide points recorded in the field. In addition, we also examined the commission and omission errors. The commission error refers to the percentage of misidentified landslides in Google Earth (100 m away from the landslide recorded in the field). The omission error refers to the percentage of landslides that were recorded in the field, but not identified in Google Earth.

2.4.4. Final Inventory Map Production

The final inventory map is the combined landslides mapped from both the field and in Google Earth. Some landslides identified in both Google Earth and field mapping were removed using the Select by Location tool in ArcGIS. In final map, the feature type of the landslides was point.

2.5. Results

Visual interpretation of Google earth imagery identified 230 landslides that occurred between 2003 to 2016 (Fig. 2.6.a). In the field, we recorded 414 landslides. We also included 57 landslides in CMA provided by Rahman et al. (2016) and 77 landslides in Cox's Bazar and Teknaf municipalities provided by CDMP-II (2012). In total, we collected 548 landslides based on field mapping (Fig. 2.6.b). The field mapping covered accessible areas where landslides were reported, and the field-recorded landslides that mainly occurred in June 2017 (356 out of 548). Among these recent landslides, 305 of them occurred in the landslide prone areas mentioned in newspaper reports and 51 occurred in new areas. In Bandarban, 101 landslides were identified and from them 25 landslides occurred before June 2017 and the oldest on dated back to 1993. In Rangamati, all field-mapped landslides occurred during June 2017. Among 82 landslides in Khagrachari, only 12 of them occurred before 2017. Out of 137 field-mapped landslides in Chittagong, 74 landslides occurred before 2017. Table 2 shows the distribution of the landslides identified in Google Earth and field mapping in the CHA. The mean elevation of landslides identified in Google Earth is 127.3 m (SD= 121.0 m), the maximum elevation is 652.0 m and 85% of the landslides are less than 200.0 m (Fig. 2.7.a). For landslides identified in the field mapping, the mean elevation is 72.0 m (SD=121.0 m), the maximum elevation is 483.0 m, and about 82% of the landslides are less than 100.0 m (Fig. 2.7.a). Identifying the type of landslide is difficult in Google Earth, depending on the quality of the imagery and the skill of the interpreter. Among 230 landslides mapped in Google Earth, 62 landslides were undefined due to the difficulty to determine their types. Slide is the dominant type of landslides and flow and fall are other two major types identified in Google Earth (Fig. 2.7.b). Among 15 unrecognized landslides, 12 are from Cox's Bazar district because we did not carry out the fieldwork there and landslide locations were provided by CDMP-II (2012). In field mapping, flow is the dominant type of landslides, accounting for 40% of the total landslides

(Fig. 2.7.b). Slide is the second dominant type. There are also 28 complex landslides which are combination of two or more types of landslides. Most field mapped landslides are shallow landslides (depth less than 10 m) with only 20 (out of 548) deep landslides which were large slope failures and depths were >10m.

The final landslide inventory includes the locations of 730 landslides, as well as their types and time of occurrence. About 48.8% (356 out of 730) of the landslides in the inventory are recent landslides. This dataset is the largest landslide inventory of the CHA (Fig. 2.6.c). The mapped landslides are clustered in some specific areas (Fig. 2.6.a-c). The clusters are associated with the natural factors that influence landslides and the areas covered during field mapping. For example, landslides are clustered near the fault lines and in areas where the slope gradient is between 10-30°. Our field mapping was mainly in urban areas and along the roads.

The validation and accuracy assessment were conducted in three test sites in Bandarban, CMA, and Cox's Bazar municipality. In our test site in Bandarban, we identified 25 landslides during field mapping and 22 landslides in Google Earth. All these landslides are <100 m buffer distance from the landslides identified in the field mapping (8 have 0 distance) (Table 2.3). Therefore, the overall percentage accuracy is 88% using the 100 m threshold buffer. The commission error is 0%, indicating that all landslides identified in Google Earth are actual landslides (Table 2.3). The omission error is 12%, meaning that 12% of the landslides we identified in the field were not mapped in Google Earth.

In CMA, we mapped 63 landslides in Google Earth. We used 44 landslides field-mapped by Rahman et al. (2016) for validation. Among the 63 landslides, 9 landslides are at 0 m distance, 30 landslides are <100 m buffer distance from the field-mapped landslides (Table 2.3). The overall

percentage accuracy is 68.2% for the 100 m threshold buffer. The commission error is 52.4 % and the omission error is 31.8% (Table 2.4). In Cox's Bazar municipality, we identified 54 landslides in Google Earth and used 64 landslides identified by field mapping of CDMP-II (2012) for validation. Among these 54 landslides, 7 landslides are at 0 m distance, 44 landslides are <100 m buffer distance from the landslide identified in the field mapping (Table 2.3). The overall percentage accuracy is 68.7% for 100 m threshold buffer. Here commission error is 18.5% and omission error is 31.3% (Table 2.4).

The apparently higher accuracy in our test site than two other areas is likely caused by the different field mapping methods used in these sites. We mapped landslides in Google Earth and validated all these landslides at the test site in Bandarban. The field-mapped landslides in CMA and Cox's Bazar municipality are likely only those causing casualties. In Cox's Bazar, field map of CDMP-II (2012) helped include landslides in high-density urban areas, but we could not identify them in Google Earth. Although Google Earth has high-resolution imageries for urban areas, it may not be enough to detect landslides in high-density urban areas. Therefore, field mapping is still the best option to detect landslides in high-density urban areas. The omission errors range from 12% to 31% in these three sites, indicating that we may miss 10-30% of the landslides in Google Earth, especially in urban areas.

2.6. Discussion and Conclusions

We produced a landslide inventory map of CHA in Bangladesh based on Google Earth imagery and field mapping. In our study, field mapping helped identify more recent landslides that occurred in June 2017 in five districts of the study area. In Bangladesh, vegetation regrows very quickly and in urban areas the rate of anthropogenic activities is very high, so the sign of landslides may be removed quickly. As our field work was conducted just one month after the occurrence of the

new landslides, we mapped more landslides in the field than in Google Earth. In addition, uncertainties and biases may exist in mapping landslides in Google Earth. Historical Google Earth imagery may not have continuity. Specifically, there is no regular monthly or yearly interval among two historical images and the time gap between two historical images can be up to several years. Landslides may occur within such large time map but cannot be included in the inventory because the vegetation regrows quickly, and the sign of landslide may not be found in the next available image. Thus, the inventory prepared by Google Earth may not be a complete one. Field mapping may help in this regard, but in our study, field mapping mostly captured landslides that occurred during June 2017. Thus, field mapping could not help reducing the uncertainty caused by the unavailability of historical imagery. Uncertainties and biases exist in the field mapping as well. We mapped landslides mainly along roads, in urban areas, and in areas where are accessible, whereas the inaccessible forest and mountain areas are excluded in the field mapping.

We developed a set of criteria to identify landslides using Google Earth imagery. These criteria can be adopted in other areas, especially in developing countries where high-resolution satellite imagery and aerial photographs are not available. We also introduced a method for differentiating areas under jhum cultivation from landslides. It can help landslide detection in areas where slash and burning are practiced. A method for accuracy assessment was developed when landslides are mapped as points rather than polygons in the field. Detail topographic maps are not available for some areas especially in developing countries and landslides polygons cannot be drawn around the landslides in the field. Our assessment method would be helpful for this type of scenarios.

This work produced an updated landslide inventory of CHA. Previous studies mainly covered three urban areas and we expanded the mapping to all districts in CHA. We found that the Rangamati district has the second highest number of landslides although few studies were

conducted there. We mapped 211 landslides in Bandarban and Khagrachari districts, accounting about 27% of the total of CHA. Therefore, this work helps refine the spatial distribution of landslides. Future work is needed to conduct the morphometric and engineering analysis on the landslides in these new areas.

References

Ahmed, B., 2014. Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides, 12*(6), pp.1077-1095.

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), pp.1707-1737.

Ahmed, B. and Dewan, A., 2017. Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in Chittagong City Corporation, Bangladesh. *Remote Sensing*, *9*(4), p.304.

Ahmed, B., Rahman, M., Islam, R., Sammonds, P., Zhou, C., Uddin, K. and Al-Hussaini, T.M., 2018. Developing a dynamic web-gis based landslide early warning system for the chittagong metropolitan area, Bangladesh. *ISPRS International Journal of Geo-Information*, *7*(12), p.485.

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C. and Pradhan, B., 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Computers & Geosciences*, *38*(1), pp.23-34.

Aleotti, P. and Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, *58*(1), pp.21-44.

Alkevli, T. and Ercanoglu, M., 2011. Assessment of ASTER satellite images in landslide inventory mapping: Yenice-Gökçebey (Western Black Sea region, Turkey). *Bulletin of Engineering Geology and the Environment*, *70*(4), pp.607-617.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides*, *11*(6), pp.1063-1078.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena, 114*, pp.21-36.

Althuwaynee, O.F., Pradhan, B. and Lee, S., 2016. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. *International Journal of Remote Sensing*, *37*(5), pp.1190-1209.

Arora, M.K., Das Gupta, A.S. and Gupta, R.P., 2004. An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. International Journal of Remote Sensing, 25(3), pp.559-572.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, *65*(1-2), pp.15-31.

Banglapedi., 2015. Landslide, Banglapedia: National Encyclopaedia of Bangladesh. Asiatic Society of Bangladesh. http://en.banglapedia.org/index.php?title Landslide. Accessed 10 March 2018

Brabb, E.E., 1991. The world landslide problem. *Episodes Journal of International Geoscience*, *14*(1), pp.52-61.

Brammer, H., 2012. *Physical geography of Bangladesh*. The University Press Ltd.

Bruschi, V.M., Bonachea, J., Remondo, J., Gómez-Arozamena, J., Rivas, V., Barbieri, M., Capocchi, S., Soldati, M. and Cendrero, A., 2013. Land management versus natural factors in land instability: some examples in northern Spain. *Environmental management, 52*(2), pp.398-416.

Budimir, M.E.A., Atkinson, P.M. and Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. *Landslides*, *12*(3), pp.419-436.

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I. and Dick, O.B., 2012. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. *Geomorphology*, *171*, pp.12-29.

Carrara, A., Guzzetti, F., Cardinali, M. and Reichenbach, P., 1999. Use of GIS technology in the prediction and monitoring of landslide hazard. *Natural hazards*, *20*(2), pp.117-135.

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, *13*(11), pp.2815-2831.

CDMP, I., 2012. Landslide Inventory and Landuse Mapping. *DEM Preparation, Precipitation Threshold Value and Establishment of Early Warning Device*.

Chen, W., Zhang, S., Li, R. and Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Science of the total environment*, *644*, pp.1006-1018.

Chen, W., Shahabi, H., Shirzadi, A., Hong, H., Akgun, A., Tian, Y., Liu, J., Zhu, A.X. and Li, S., 2019. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment, 78(6), pp.4397-4419.

Cruden, D.M. and Varnes, D.J., 1996. Landslides investigation and mitigation. Landslide types and processes. *Special report*, *247*.

Crosby, C.J., 2012. Lidar and Google Earth: Simplifying access to high-resolution topography data. *Geological Society of America Special Papers*, *492*, pp.37-47.

Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y. and Zhu, Z., 2015. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Natural Hazards, 78(3), pp.1749-1776.

Đurić, D., Mladenović, A., Pešić-Georgiadis, M., Marjanović, M. and Abolmasov, B., 2017. Using multiresolution and multitemporal satellite data for post-disaster landslide inventory in the Republic of Serbia. *Landslides*, *14*(4), pp.1467-1482.

Ferdous, L., Kafy, A.A., Roy, S. and Chakma, R., 2020, August. An analysis of Causes, Impacts and Vulnerability Assessment for Landslides Risk in Rangamati District, Bangladesh. In *International Conference on Disaster Risk Mitigation., Dhaka, Bangladesh. http://www. bip. org. bd/SharingFiles/201805080735531. pdf. Accessed* (Vol. 11).

Fisher, G.B., Amos, C.B., Bookhagen, B., Burbank, D.W. and Godard, V., 2012. Channel widths, landslides, faults, and beyond: The new world order of high-spatial resolution Google Earth imagery in the study of earth surface processes. *Geological Society of America Special Papers*, *492*(01), pp.1-22.

Frattini, P., Crosta, G. and Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. *Engineering geology*, *111*(1-4), pp.62-72.

Felicísimo, Á.M., Cuartero, A., Remondo, J. and Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*, *10*(2), pp.175-189.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology*, *94*(3-4), pp.268-289.

Glade, T., 2001. Landslide hazard assessment and historical landslide data—an inseparable couple?. In *The use of historical data in natural hazard assessments* (pp. 153-168). Springer, Dordrecht.

Gomez, H. and Kavzoglu, T., 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Engineering Geology*, *78*(1-2), pp.11-27.

Guzzetti, F., 2002, October. Landslide hazard assessment and risk evaluation: Limits and prospectives. In *Proceedings of the 4th EGS Plinius Conference, Mallorca, Spain* (pp. 2-4).

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M., 2006. Estimating the quality of landslide susceptibility models. *Geomorphology*, *81*(1-2), pp.166-184.

Guzzetti, F., Ardizzone, F., Cardinali, M., Rossi, M. and Valigi, D., 2009. Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth and Planetary Science Letters*, *279*(3-4), pp.222-229.

Sifa, S.F., Haque, D.M.E., Mahmud, T. and Tarin, M.A., 2018, December. Landslide Susceptibility Assessment Based on Satellite Image Processing and Bi-variate Statistical Modeling for Rangamati District, Bangladesh. In *AGU Fall Meeting 2018*. AGU.

Heckmann, T., Gegg, K., Gegg, A. and Becht, M., 2014. Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. Natural Hazards and Earth System Sciences, 14(2), p.259.

Hölbling, D., Eisank, C., Albrecht, F., Vecchiotti, F., Friedl, B., Weinke, E. and Kociu, A., 2017. Comparing manual and semi-automated landslide mapping based on optical satellite images from different sensors. *Geosciences*, *7*(2), p.37.

Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W. and Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, pp.399-413.

Huabin, W., Gangjun, L., Weiya, X. and Gonghui, W., 2005. GIS-based landslide hazard assessment: an overview. *Progress in Physical geography*, *29*(4), pp.548-567.

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C. and Bălteanu, D., 2016. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology*, *253*, pp.508-523.

Jiménez-Perálvarez, J.D., El Hamdouni, R., Palenzuela, J.A., Irigaray, C. and Chacón, J., 2017. Landslide-hazard mapping through multi-technique activity assessment: an example from the Betic Cordillera (southern Spain). *Landslides*, *14*(6), pp.1975-1991.

Kanwal, S., Atif, S. and Shafiq, M., 2017. GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomatics, Natural Hazards and Risk*, *8*(2), pp.348-366.

Khan, Y.A., Lateh, H., Baten, M.A. and Kamil, A.A., 2012. Critical antecedent rainfall conditions for shallow landslides in Chittagong City of Bangladesh. *Environmental Earth Sciences*, *67*(1), pp.97-106.

Liao, Z., Hong, Y., Kirschbaum, D. and Liu, C., 2012. Assessment of shallow landslides from Hurricane Mitch in central America using a physically based model. *Environmental Earth Sciences*, *66*(6), pp.1697-1705.

Li, C., Yan, J., Wu, J., Lei, G., Wang, L. and Zhang, Y., 2019. Determination of the embedded length of stabilizing piles in colluvial landslides with upper hard and lower weak bedrock based on the deformation control principle. *Bulletin of engineering geology and the environment*, *78*(2), pp.1189-1208.

Lombardo, L. and Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results. *Engineering geology*, *244*, pp.14-24.

Marchesini, I., Ardizzone, F., Alvioli, M., Rossi, M. and Guzzetti, F., 2014. Non-susceptible landslide areas in Italy and in the Mediterranean region. *Natural Hazards and Earth System Sciences*, *14*(8), pp.2215-2231.

Masum, K.M., Rashid, M.M., Jashimuddin, M. and Ara, S.J.G., 2011. Land use conflicts of Chittagong Hill tracts and indigenous hill people as victim in Bangladesh. *J. Gen. Educ*, *1*, pp.62-71.

Mondini, A.C., Chang, K.T. and Yin, H.Y., 2011. Combining multiple change detection indices for mapping landslides triggered by typhoons. *Geomorphology*, *134*(3-4), pp.440-451.

Myronidis, D., Papageorgiou, C. and Theophanous, S., 2016. Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP). *Natural Hazards*, *81*(1), pp.245-263.

Othman, A.A., Gloaguen, R., Andreani, L. and Rahnama, M., 2018. Improving landslide susceptibility mapping using morphometric features in the Mawat area, Kurdistan Region, NE Iraq: Comparison of different statistical models. *Geomorphology*, *319*, pp.147-160.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T., 2014. Assessing the quality of landslide susceptibility maps–case study Lower Austria. *Natural Hazards and Earth System Sciences*, *14*(1), pp.95-118.

Prothom Alo ., 2017. Rangamati landslide death toll 118. Available at: https://en.prothomalo.cpm/bangladesh/news/151605/Rangamati-Landslide -death toll-118.

Rabby, Y.W. and Li, Y., 2019. An integrated approach to map landslides in Chittagong Hilly Areas, Bangladesh, using Google Earth and field mapping. *Landslides*, *16*(3), pp.633-645.

Rahman, M.S., Rahman, B.A.F.H.S. and T, M., 2016. Landslide inventory in an urban setting in the context of Chittagong Metropolitan Area, Bangladesh.

Rahman, M.S., Ahmed, B. and Di, L., 2017. Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science*, *14*(10), pp.1919-1937.

Rasyid, A.R., Bhandary, N.P. and Yatabe, R., 2016. Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. *Geoenvironmental Disasters*, *3*(1), pp.1-16.

Reichenbach, P., Mondini, A.C. and Rossi, M., 2014. The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). *Environmental management*, *54*(6), pp.1372-1384.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, pp.60-91.

Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T. and Akgun, A., 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arabian Journal of Geosciences*, *7*(2), pp.725-742.

Remondo, J., González, A., De Terán, J.R.D., Cendrero, A., Fabbri, A. and Chung, C.J.F., 2003. Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain. *Natural Hazards*, *30*(3), pp.437-449.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C. and Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology*, *114*(3), pp.129-142.

Roy, J. and Saha, S., 2019. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenvironmental Disasters*, *6*(1), pp.1-18.

Sabokbar, H.F., Roodposhti, M.S. and Tazik, E., 2014. Landslide susceptibility mapping using geographically-weighted principal component analysis. *Geomorphology*, *226*, pp.15-24.

Sabatakakis, N., Koukis, G., Vassiliades, E. and Lainas, S., 2013. Landslide susceptibility zonation in Greece. *Natural hazards*, *65*(1), pp.523-543.

Samodra, G., Chen, G., Sartohadi, J. and Kasama, K., 2018. Generating landslide inventory by participatory mapping: an example in Purwosari Area, Yogyakarta, Java. *Geomorphology*, *306*, pp.306-313.

Sato, H.P. and Harp, E.L., 2009. Interpretation of earthquake-induced landslides triggered by the 12 May 2008, M7. 9 Wenchuan earthquake in the Beichuan area, Sichuan Province, China using satellite imagery and Google Earth. *Landslides*, *6*(2), pp.153-159.

Schicker, R. and Moon, V., 2012. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology*, *161*, pp.40-57.

Segoni, S., Rosi, A., Lagomarsino, D., Fanti, R. and Casagli, N., 2018. Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Natural Hazards and Earth System Sciences*, *18*(3), pp.807-812.

Sifa, S.F., Mahmud, T., Tarin, M.A. and Haque, D.M.E., 2020. Event-based landslide susceptibility mapping using weights of evidence (WoE) and modified frequency ratio (MFR) model: A case study of Rangamati district in Bangladesh. *Geology, Ecology, and Landscapes*, *4*(3), pp.222-235.

Shirzadi, A., Shahabi, H., Chapi, K., Bui, D.T., Pham, B.T., Shahedi, K. and Ahmad, B.B., 2017. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. Catena, 157, pp.213-226.

Steger, S., Brenning, A., Bell, R., Petschko, H. and Glade, T., 2016. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology*, *262*, pp.8-23.

Steckler, M.S., Nooner, S.L., Akhter, S.H., Chowdhury, S.K., Bettadpur, S., Seeber, L. and Kogan, M.G., 2010. Modeling Earth deformation from monsoonal flooding in Bangladesh using hydrographic, GPS, and Gravity Recovery and Climate Experiment (GRACE) data. *Journal of Geophysical Research: Solid Earth*, *115*(B8).

Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M. and Sorichetta, A., 2011. Spatial agreement of predicted patterns in landslide susceptibility maps. *Geomorphology*, *125*(1), pp.51-61.

Thanh, L.N. and De Smedt, F., 2012. Application of an analytical hierarchical process approach for landslide susceptibility mapping in A Luoi district, Thua Thien Hue Province, Vietnam. *Environmental Earth Sciences*, *66*(7), pp.1739-1752.

Tsangaratos, P. and Benardos, A., 2014. Estimating landslide susceptibility through a artificial neural network classifier. *Natural Hazards*, *74*(3), pp.1489-1516.

Vakhshoori, V. and Zare, M., 2016. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomatics, Natural Hazards and Risk*, *7*(5), pp.1731-1752.

Wang, Q., Wang, Y., Niu, R. and Peng, L., 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*, *9*(9), p.938.

Xu, L., Coop, M.R., Zhang, M. and Wang, G., 2018. The mechanics of a saturated silty loess and implications for landslides. *Engineering Geology*, *236*, pp.29-42.

Yan, F., Zhang, Q., Ye, S. and Ren, B., 2019. A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model. *Geomorphology*, *327*, pp.170-187.

Yang, K., Sun, L.P., Huang, Y.X., Yang, G.J., Wu, F., Hang, D.R., Li, W., Zhang, J.F., Liang, Y.S. and Zhou, X.N., 2012. A real-time platform for monitoring schistosomiasis transmission supported by Google Earth and a web-based geographical information system. *Geospatial health*, pp.195-203.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences*, *35*(6), pp.1125-1138.

Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environmental Earth Sciences*, *61*(4), pp.821-836.

Zêzere, J.L., Pereira, S., Melo, R., Oliveira, S.C. and Garcia, R.A., 2017. Mapping landslide susceptibility using data-driven methods. *Science of the total environment*, *589*, pp.250-267.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T. and Hong, H., 2019. A similarity-based approach to sampling absence-data for landslide susceptibility mapping using data-driven methods. *Catena*, *183*, p.104188.

Appendix

Table 2.1. List of Main Landslide Information Sources

Source of Information	Information Collected
Local Newspaper	Date, time and locations of landslides
Rahman et al. (2016)	57 landslide locations of CMA
CDMP-II (2012)	77 landslide locations of Cox's Bazar and Teknaf municipalities
Records of Department of Disaster Management of People's Republic of Bangladesh	Name of the locations of landslide that caused casualties
Roads and Highways Department	Locations of landslides that caused road damages

Table 2.2. Distribution of Landslides identified in Google Earth and Field Mapping among Districts of Chittagong Hilly Areas

Districts	Google Earth	Field Mapped
Chittagong	121	137
Bandarban	22	101
Cox's Bazar	48	77
Khagrachari	6	82
Rangamati	33	151

Table 2.3. Percentage of Landslide Locations at different Distance from Ground Points in Bandarban, CMA and Cox's Bazar

Study Site	Distance (m)	0	1-10	10-20	20-50	50-100	Above 100
Bandarban	Number of Landslides	8	1	3	4	6	3
	Percentage	32.0	4.0	12.0	16.0	24.0	12.0
	Cumulative Percentage	32.0	36.0	48.0	64.0	88.0	100.0
CMA	Number of Landslides	9	5	3	5	8	14
	Percentage	20.5	11.6	6.8	11.4	18.2	31.9
	Cumulative Percentage	20.5	31.8	38.6	50.0	68.2	100.0
Cox's Bazar	Number of Landslides	7	18	8	8	3	20
	Percentage	10.9	28.1	12.5	12.5	4.7	31.3
	Cumulative Percentage	10.9	39.0	51.5	64.0	68.7	100.0

Table 2.4. Accuracy Assessment Table for Bandarban, CMA and Cox's Bazar (Column: Field mapping Row: Google Earth)

Study Sites		Landslide	Non-Landslide	Total	Commission Error (%)
Bandarban	Landslide	22	0	22	0.0
	Non-Landslide	3	-		
	Total	25			
	Omission Error (%)	12.0			
CMA	Landslide	30	33	63	52.39
	Non-Landslide	14	-		
	Total	44			
	Omission Error (%)	31.82			
Cox's Bazar	Landslide	44	10	54	19.52
	Non-Landslide	20	-		
	Total	64			
	Producer's Accuracy (%)	68.75			
	Omission Error (%)	31.25			

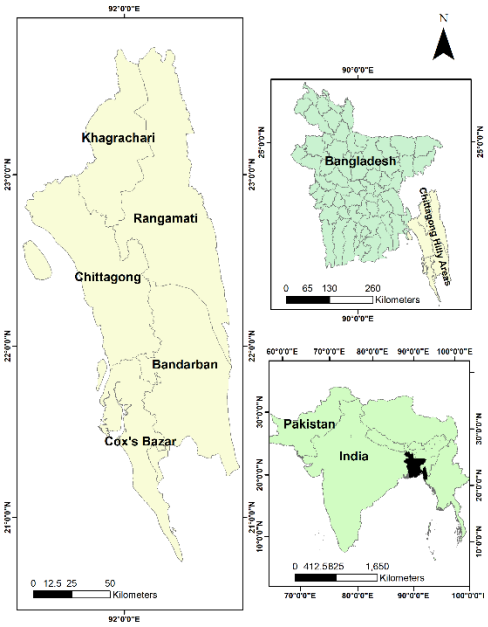

Fig. 2.1. Geographical Position of Chittagong Hilly Areas

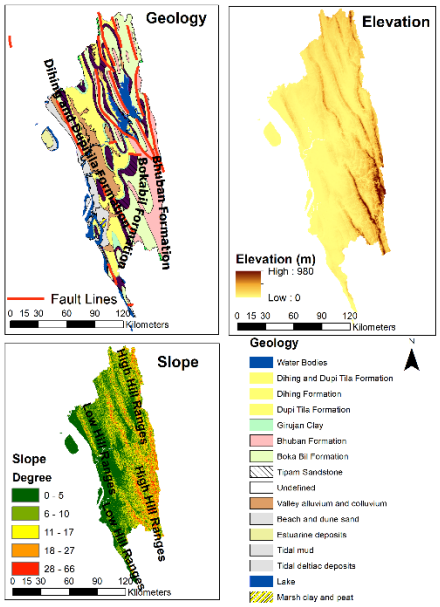

Fig. 2.2. Geological, Slope, and Elevation Maps of Chittagong Hilly Areas

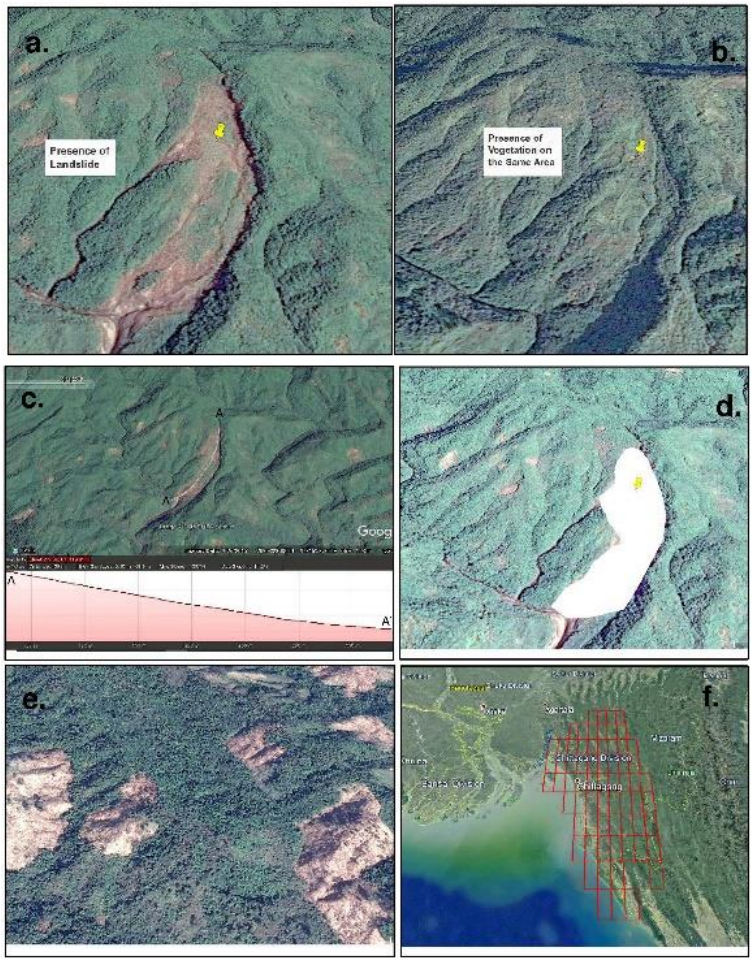

Fig. 2.3. Landslide Detection in Google Earth. (a) and (b): Change Detection and Identification in Google Earth; (c): Landslide Identification through Elevation Profile in Google Earth; and (d): Polygon Drawn around the Scarp and Run out of Landslide (e) Presence of Clear-Cut (f) Fishnet

Fig. 2.4. Field Mapping. (a) and (b): Field Mapping with the Assistance of Local People, (c) to (f): Identification of Landslides and GPS Coordinate Collection

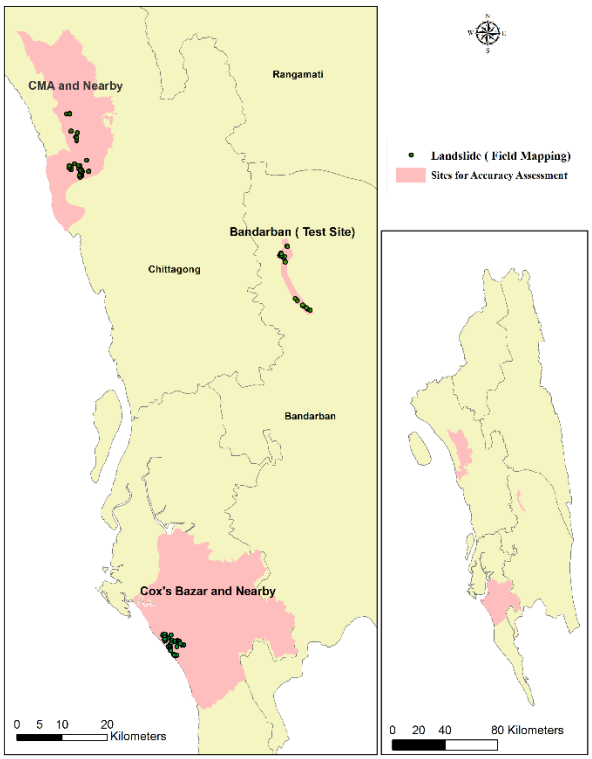

Fig. 2.5. Location of Study Sites for Map Validation and Accuracy Assessment

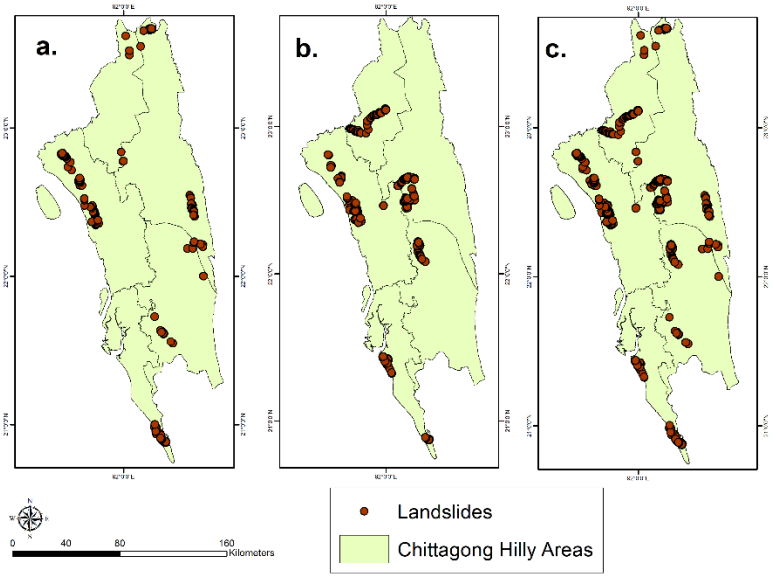

Fig. 2.6. Landslide Inventory Maps of Chittagong Hilly Areas of Bangladesh. (a): Landslide Inventory Map based on Google Earth; (b): Landslide Inventory Map based on Field Mapping; and (c) Final Landslide Inventory Map

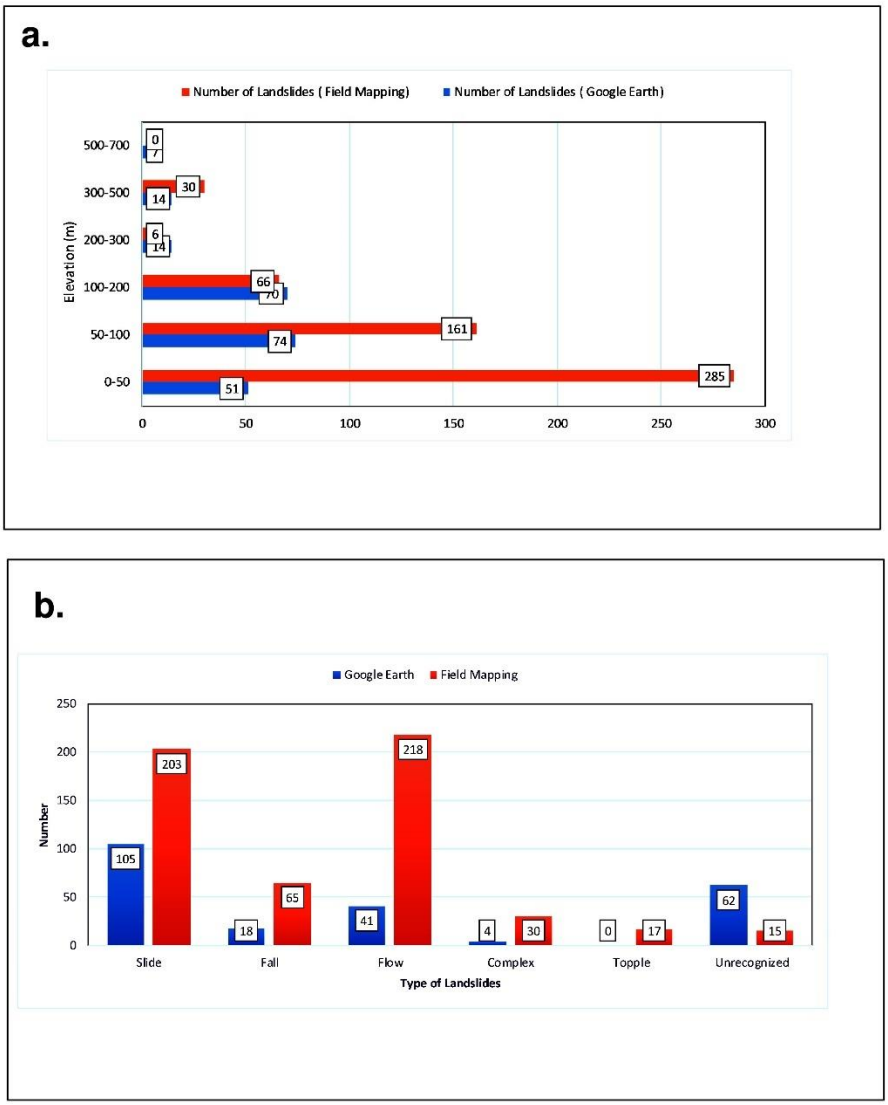

Fig. 2.7. Different Statistics of Identified Landslides in Google Earth and Field Mapping. (a): Number of Landslides at different Elevation (Google Earth and Field Mapping) based on ASTER 30 m DEM; and (b) Number of Different Types of Landslides (Google Earth and Field Mapping)

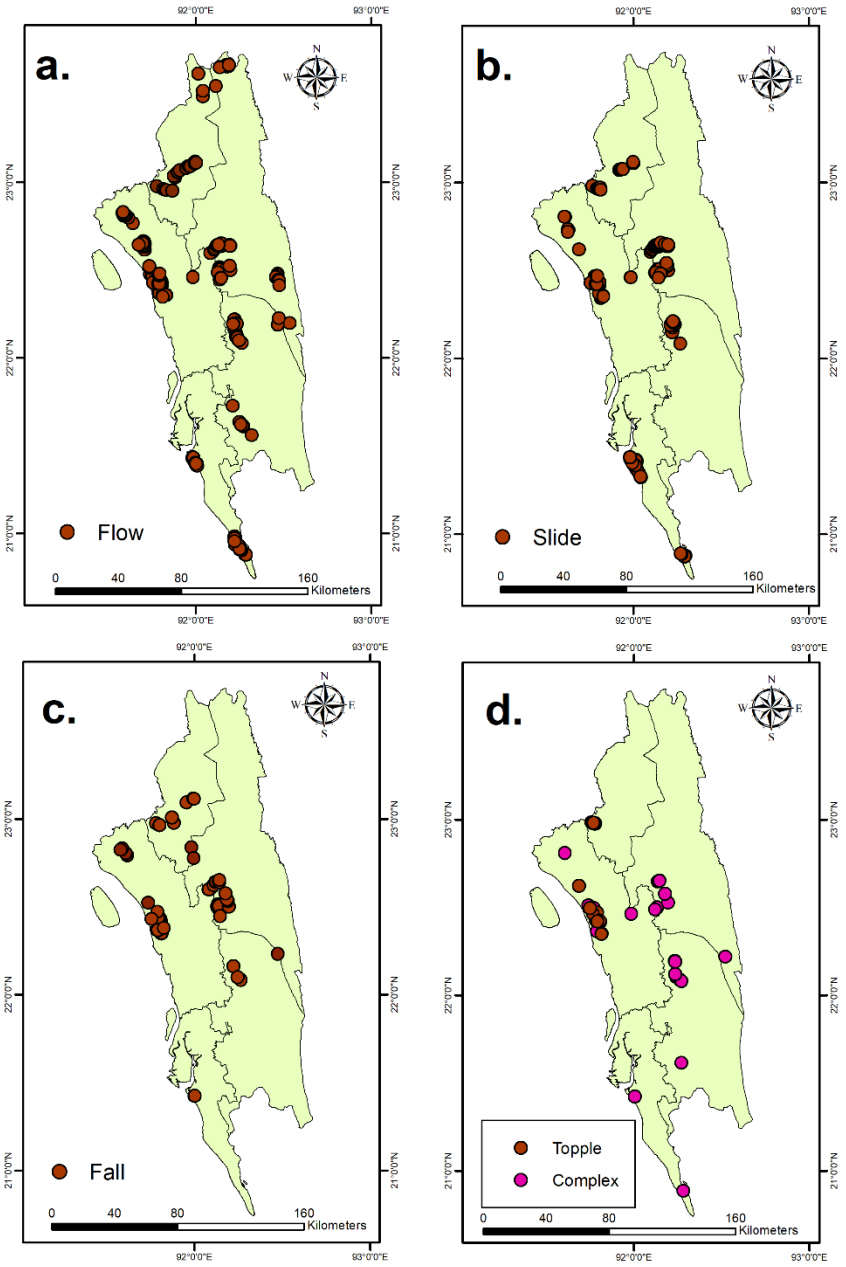

Fig. 2.8. Distribution of Different Types of Landslides in Chittagong Hilly Areas of Bangladesh. (a): Slide; (b): Flow (c) Fall and (d) Topple and Complex

Chapter 3

An objective method to determine absence-data sampling for landslide susceptibility mapping

This chapter is a manuscript in preparation for *Landslide*. The use of "we" in this chapter refers to co-authors, Drs. Yingkui Li, Haileab Hilafu, and me. As the first author, I conducted the field work, led the data analysis, and wrote the manuscript.

Abstract

The accuracy and quality of the landslide susceptibility map depend on the available landslide locations and the sampling strategy for absence-data (non-landslide locations). In this study, we proposed an objective method to determine the critical value for sampling absence-data based on the Chi-square distribution of the Mahalanobis distances (MD) and a user-specified confidence level. We demonstrated this method on landslide susceptibility mapping of three subdistricts (Upazilas) of the Rangamati district, Bangladesh, and compared the results with the landslide susceptibility map produced based on a widely used slope-based absence-data sampling method. We first determined the critical value of 23.69 (at 95% confidence level) based on the Chi-square distribution of the MD values of the 261 landslide locations derived using 15 landslide causal factors, including slope, aspect, plan curvature, and profile curvature. This critical value was then used to determine the sampling space for 261 random absence-data. In comparison, we chose another set of the absence-data based on a slope threshold of <3º. The landslide susceptibility maps were then generated using the random forest model based on the landslide and non-landslide samples. The success and prediction rates and the Kappa index were used for accuracy assessment, while the Seed Cell Area Index (SCAI) was used for consistency assessment. Landslide susceptibility map produced using our proposed method has relatively high success (88.4%), prediction (86.2%), and Kappa values (0.75). The SCAI values also indicate that the landslide susceptibility map is consistent. In contrast, even though the landslide susceptibility map produced by the slope-based sampling also has relatively high accuracy, the SCAI values suggest lower consistency. Furthermore, the slope-based sampling is highly subjective; therefore, we recommend using the MD-based absence-data sampling for landslide susceptibility mapping.

3.1. Introduction

Landslides are the movement of rock, soil, and earth along a slope (Cruden and Varnes, 1992) when the shear stress on the slope materials exceeds the shear strength (Ahmed and Dewan, 2017). It causes damage to infrastructure and the loss of human lives worldwide (Guzzetti et al., 2002; Yilmaz, 2009; Yilmaz, 2010). Landslide inventory and susceptibility mapping are critical to mitigate the losses caused by landslides (Ahmed, 2015; Ahmed and Dewan, 2017; Chen et al., 2017; Hong et al., 2018; Ahmed et al., 2018). Landslide inventory documents previously occurred landslides (Guzzetti et al. 2012), while landslide susceptibility describes the probability of landslides over an area (Sterlachini et al., 2011). Landslides are affected by various causal factors, such as slope, curvature, land use/land cover, geology, and elevation (Althuwaynee et al., 2014; Althuwaynee et al. 2016; Chen et al. 2017). Landslide inventory and its relationship with different causal factors can be used to derive the landslide susceptibility map (Reichenbach et al., 2020). Various statistical methods have been used for landslide susceptibility mapping, including logistic regression, support vector machines, random forest, and gradient boosting (Ayalew and Yamagishi, 2005, Vakshoori and Zare, 2016; Reicehnbach et al. 2018). These statistical methods use landslide causal factors as independent variables and landslide locations (presence data) and non-landslide locations (absence-data) as dependent variables (Yilmaz, 2009). The presence data are mainly from the landslide inventory. In contrast, the absence-data are usually unavailable and require a specific sampling strategy (Zhu et al. 2014; Chen et al. 2017). The quality and accuracy of the landslide susceptibility maps depend not only on the quality of causal factors and presence data but also on the sampling strategy for the absence-data (Zhu et al., 2017).

Random sampling is the most common approach for the absence-data. It considers all locations other than the recorded landslides for absence-data (Tsangaratos and Benardos, 2014; Regmi et al., 2014). This method requires a representative landslide inventory of all landslides for the whole area (Zhu et al., 2019). It is suitable for the landslide susceptibility mapping in a relatively small area but faces challenges at a large area or regional scale (Althuwaynee et al. 2014). The accuracy of the landslide susceptibility map based on random sampling is generally low and biased towards the known landslide locations (Zhu et al., 2019). Various absence-data sampling methods have been proposed to improve the accuracy and quality of landslide susceptibility mapping, including prior data exploratory analysis, buffer-controlled sampling, distance and density-based measures like Kernel density estimation, Euclidean distance, one class or presence only classification

method, and species density distribution modeling like Bioclim (Althuwaynee et al. 2014; Chen et al. 2017; Hong et al. 2018; Zhu et al. 2019).

Prior data exploratory analysis determines a safe zone for absence-data sampling based on the available landslide locations (Chen et al., 2017; Huang et al., 2017; Adnan et al., 2020). This method generally chooses one of the most important causal factors, such as slope and geology, to determine the safe zone for the absence-data sampling (Althuwaynee et al. 2014, Huang et al., 2017). However, the results generated using this method are biased towards the selected factor. For instance, if the safe zone is determined based on slope, the model will likely be biased towards the slope (Hong et al., 2018). Yao et al. (2008) used a buffer-controlled sampling method, assuming that the areas near each other are more similar than those distant apart. The selection of the buffer distance is subjective because it depends on expert knowledge (Zhu et al., 2019). Hong et al. (2018) proposed a one-class classification or presence only method similar to the one-class support vector machine method. In this method, classification like absence and presence data are not given in the model's training stage. Only the presence data is used to classify an area into two parts: one part is similar to the presence data or landslides, and the other has dissimilarities with the landslides. The area with high dissimilarities is used for absence-data sampling.

Distance-based sampling assumes that areas with similar environmental conditions (explained by the causal factors) experience similar geomorphic processes like landslides (Zhu et al., 2019; Hong et al., 2018). A distance threshold, known as the critical value, is needed to determine the sampling space for absence-data (Tsangaratos and Benardos, 2014). Although several distance-based measures have been used, determining this critical value is not explained (Zhu et al., 2019). Generally, users select the critical value subjectively to maximize the accuracy of the landslide susceptibility map (Hong et al., 2018). Moreover, only one method, like the area under the curve or Continuous Boyce Index, is used to assess the mapping accuracy (Zhu et al. 2014; Zhu et al. 2019; Reichenbach et al., 2018) without the assessment of the mapping consistency (Reichenbach et al., 2018; Abedini and Tulabi, 2018). A landslide susceptibility model can achieve high accuracy by increasing the area under high and very high landslide-prone zones. However, it may overestimate the landslide susceptibility by assigning landslide-free areas as prone zones (Schiker and Moon, 2012). It is impossible to implement the overestimated map for practical purposes as the map loses its consistency (Reichenbach et al., 2018). Zhu et al. (2019) found that decreasing the sampling space of the absence-data increases the accuracy of the landslide susceptibility map,

but may overestimate the landslide susceptibility (Hong et al., 2018; Zhu et al., 2019). It is essential to choose the critical value or threshold for to satisfy both accuracy and consistency.

In this work, we proposed an objective method to determine the critical value of absence-data sampling based on the Chi-square distribution of the Mahalanobis distance and a user-specified confidence level. We applied this proposed method to the landslide susceptibility mapping in the three Upazilas (sub-district) of the Rangamati district, Bangladesh, and compared the model performance with a traditionally used slope-based method for absence-data sampling.

3.2. Methodology

This study employed the third law of geography to determine sampling space for absence-data sampling. According to the third law of geography, if two areas have the same geographical environment, they will experience the same geographical processes such as landslides (Zhu et al., 2019). The characteristics of the geographic environment used in this study are the landslide causal factors. Since we are searching for sampling space for (landslide) absence-data sampling, we must find out areas with the least similarities with landslide locations. We assume that landslide locations will have a geomorphic environment defined by the landslide causal factors. For example, the slope is a landslide causal factor, and for all the landslide locations, there will be a typical value of slope (e.g., the average slope for the observed landslide locations). We seek locations whose slope possesses the highest dissimilarities with the typical slope of the landslide locations. If we have n number of landslide locations and p number of causal factors, then these locations will have a mean environmental condition based on the p causal factors. Non-landslide locations will be farther away from that mean condition. This study employs Mahalanobis distance to measure the distance between the mean landslide condition and the condition of a potential site to determine whether it has similarities or dissimilarities with the landslide locations.

3.2.1. Mahalanobis Distance

Mahalanobis Distance (MD) is a distance metric that measures the distance between the data point location and the distribution of datasets (Nader et al., 2014; Prabhakaran, 2020). MD is an extension of the Euclidean Distance metric and can improve clustering and classification algorithms (Tsangaratos and Benardos, 2014). The Euclidean distance measures the distance between two points in p-dimensional space. It works well when the dimensional spaces are independent of each other (Prabhakaran, 2020). MD takes care of this interdependency of the dimensional spaces by dividing the Euclidean distance with the covariance matrix (Tsangaratos

and Benardos, 2014). More specifically, the MD of a potential (non-landslide) point represented by a vector of causal factors X from the centroid representation of a landslide point cloud with mean vector m and a covariance matrix C given by:

$$MD = \sqrt{(X-m)^T C^{-1}}(X-m) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

As can be seen from (1), MD reduces the correlation of variables by dividing the distance matrix by the covariance matrix (Nader et al., 2014). MD has been generally used in outlier detection and multi-class classifications (Prabhakaran, 2020). In landslide susceptibility mapping, MD can be used to define the sampling space for absence-data. The recorded landslide locations only cover a very small portion of the whole study area. Therefore, a large part of the area is not classified as landslides or non-landslides (Prabhakaran, 2020). Based on landslide locations and distribution of the causal factors, MD defines the similarity of an area to landslides' conditions. If the similarity is high, the area has a high chance for landslide and is not suitable for absence-data sampling.

It is, however, hard to determine if the similarity of an area is different enough for the absence of data sampling. Some studies used the 5^{th} quantile value to define the absence sampling space (Tsangaratos and Benardos, 2014). Zhu et al. (2019) tested a set of user-defined thresholds to determine the appropriate value for landslide susceptibility mapping. The work demonstrated that reducing absence sampling space continuously increases in the accuracy but overestimates the landslide susceptibility. However, this simple try-out strategy does not provide a statistical means to determine the optimal threshold value for absence-data sampling.

We proposed an approach to offer a statistical means for determining the MD threshold for absence-data sampling. The MD is a normalized quantity. If the causal factors have a distribution that the p-variate Gaussian distribution can approximate, the MD follows a Chi-squared distribution with *p-1* degrees of freedom. Even if the causal factors do not have an approximate *p*-variate Gaussian distribution, the MD has an approximate Chi-squared distribution with *p-1* degrees of freedom, as long as the number of causal factors is large enough (Nader et al., 2014). Based on this assumption, a critical value can be determined for a specified significance level, such as the commonly adopted significance level of 0.05. For example, if we use 15 causal factors in our study, the critical value of the MD, i.e., an MD beyond which we would conclude a potential non-landslide location is a viable sample, is 23.69. That is, when the MD is greater than this critical

value, it is considered as an outlier or different enough from the rest of the data (Nader, 2014). Therefore, we use such a critical value to determine the locations for absence-data sampling.

Fig. 3.1 shows the flow chart of our proposed method. As stated above, *n* represents the number of available landslide locations, and *p* represents the number of causal factors. A critical value is determined based on the *p-1* degrees of freedom. This critical value determines if a new point or location is a potential candidate for absence-data sampling. For any new candidate location, MD was calculated based on the mean value and the covariance matrix of the distribution of the causal factors of the *n* landslide locations. A location or point with an MD value greater than the critical value is designated as a safe zone for absence-data sampling.

To demonstrate the efficiency of this proposed method, we applied it to the landslide susceptibility mapping on three Upazilas (sub-district) of the Rangamati district, Bangladesh, and compare its derived landslide susceptibility map with the map produced based on a traditional slope-based method for absence-data sampling.

3.3. Case Study

3.3.1. Study Area and Landslide Inventory

This study focused on three Upazilas (sub-district) of the Rangamati district, Bangladesh: Rangamati Sadar, Kaptai, and Kawkhali (Fig 3.2). Rangamati Sadar is the largest city in this area. In June 2017, more than 100 people were killed by landslides in this district, and these three Upazilas were the most affected areas (Abedin et al., 2020). This district covers 1145 km^2 (BBS, 2011) with an elevation range from 7 to 576 m above mean sea level and a slope range from 0º to 52º. The western part of the area has a comparatively gentle slope, while the west and central regions are relatively steep. The bedrock of this area comprises several geological formations, including Dihing, Dupitila, Girujan Clay, Bhuban, Bokabil, and Tipam Sandstone (Rabby et al. 2020). Most of the area is covered by natural vegetation or plantation agricultural fields. Plantation agriculture and unplanned land use/land cover changes create conducive conditions, and intensive rainfall triggers landslides in this area (Ahmed, 2015; Abedin et al., 2020).

A total of 261 landslide locations (Fig 3.2) were recorded from January 2001 to January 2019. These landslides were collected by Rabby and Li (2020) based on the integrated field and Google Earth mapping and Rabby et al. (2020) based on Google Earth mapping.

3.3.2. Landslide Causal Factors

We used 15 landslide causal factors for landslide susceptibility mapping (Fig 3.3-3.4). The raster maps of these factors were prepared by Abedin et al. (2020) and Rabby et a. (2020). Table 3.1 lists the factors, resolutions, types, and data sources of these raster maps. Since the resolution of most factors is 30 m, we selected 30 m as the resolution for the landslide susceptibility mapping.

3.3.3. Absence-data Sampling

We derived the MD values for all landslide locations based on the 15 causal factors. MD value was ranged between 1.2 to 200.8 (Fig. 3.5). The degree of freedom for the Chi-square distribution of these 15 factors is 14, resulting in a critical value of 23.69 for the significance level of 0.05. We calculated the MD value for each location based on the mean and covariance matrix derived from the landslide locations and then applied this critical value to determine the sampling space for absence-data (Fig 3.5). Specifically, the locations where MD values are greater than the threshold are used for absence-data sampling to generate 261 absence-data randomly.

In comparison, we also used a slope-based sampling to determine the low landslide probability area for absence-data (Chen et al. 2018). The slope threshold is determined based on expert knowledge and judgment. Adnan et al. (2020) used the slope of $<2^{o}$ for absence-data sampling in Cox's Bazar district of Bangladesh. Ali et al. (2021) determined areas where slope $<3^{o}$ for absence-data sampling in their study in the Kysuca river basin of Slovakia. We used a threshold of slope$<3^{o}$ to randomly sample the 261-absence-data (Fig. 3.6).

3.3.4. Landslide Susceptibility Mapping

We used the random forest model to produce the landslide susceptibility maps. The random forest model proposed by Breiman (2001) is an ensemble learning method (James et al., 2017). Bootstrap aggregation is employed in RF to select subsets of observation. It generates a set of decision trees (Zhu et al., 2019; Rabby et al., 2020) and decorrelates the trees (James et al., 2017). The ensembles of decision trees decided the class membership of the dependent variables based on the highest number of votes (Pham et al., 2020). While training the model, instead of using all the predictors, RF uses a random sample of predictors (James et al., 2017). There can be a couple of strong predictors in a study, and in splitting the trees, these predictors will have an influence. RF uses a subset of predictors to overcome this problem (Zhu et al., 2019; Rabby et al., 2020). Since all the datasets are not used in modeling, the unused data are known as out-of-bag (OOB) (Youssef et al., 2016). These unselected datasets are used in determining the error and importance

of the predictors in the model (James et al., 2017). We used the "randomForest" package in R to develop the RF model for the landslide susceptibility mapping (Liaw and Weiner, 2001).

As described earlier, we generated the same number of non-landslide locations (261). This produced a dataset of 522 (261: presence data; 261 absence-data). We divided the dataset into training (391: 75%) and validation datasets (130:25%) for the landslide susceptibility mapping. In the MD-based sampling method, we used all 15 factors for the landslide susceptibility mapping. We did not include slope in the landslide susceptibility mapping for the slope-based method because the absence-data were sampled based on the slope threshold.

3.3.5. Evaluation of the model performance and consistency

3.3.5.1. Performance Assessment

We used success and prediction rate curves and statistical index-based measures to assess the model performance. The success and prediction rates are produced by plotting the landslide susceptibility or probability on X-axis and cumulative percentage of landslides on Y axis (Cheng and Fabbri, 1999). In order to compare the success and prediction rates we used the area under the curve (AUC) method which shows the area in terms of percentage of area under the graph (Carrara et al. 2008). The training dataset was used to calculate the area under the curve (AUC) of the success rate, and the validation dataset was used to calculate the AUC of the prediction rate. AUC values range from 0-100%. The greater the value, the better is the model. Generally, an AUC value >70% is considered as a fair model, and <50% indicates that the model is classifying the data randomly (Althuwaynee et al., 2016, Rasyid et al., 2016). The steeper is the curve the larger is the number of landlside locations fall into the most susceptible classes (Sterlacchini et al. 2011).

We also derived statistical index-based measures: true positive rate (TPR) (sensitivity), true negative rate (TNR) (specificity), and Kappa index. TPR measures the proportion of landslide pixels were classified correctly as landslide pixels by the model. TNR implies the proportion of absence-data that are correctly classified as absence-data by the model (Chen et al., 2017). Kappa index (Eq 2) is the ratio of observed and expected agreement, representing the model's reliability (Pham et al. 2020; Chen et al. 2017).

$$Kappa = \frac{P_{obs} - P_{exp}}{1 - P_{exp}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots .. (2)$$

Where

P_{obs}= observed agreement

P_{exp}= expected agreement

$$P_{obs} = \frac{TP + TN}{n} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$P_{exp} = \frac{(TP + FN)(TP + FP)(FP + TN)(FN + TN)}{\sqrt{N}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where

TP= true positive

TN=true negative

FN=false negative

FP=false positive

n= proportion of pixel that are classified correctly

N= the number of total training pixels

Kappa index ranges from 0-1 where 0 indicates the agreement occurred due to random guess. Whereas 1 indicates perfect agreement.

3.3.5.2. Consistency Assessment

The seed cell area index (SCAI) proposed by Suzen and Doyuran (2004) was used for the consistency assessment of the models. SCAI is the ratio between the areal extent of susceptibility classes and the percentage of landslides that occurred in the susceptibility classes and can be described as Eq 5.

$$SCAI = \frac{N_i}{n_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where

N_i= percentage of area under i susceptibility class

n_i= percentage of landslides under i susceptibility class

SCAI value ranges from 0 to ∞. The smaller is the SCAI value, the more consistent the model is. SCAI value decreased from a very low susceptibility zone to a very high susceptibility zone

(Arabameri et al. 2020). This index determines whether landslide locations or pixels are spread over a very conservative areal extent (Sdao et al., 2013). It can identify whether a model is overestimating landslide susceptibility. If a model overestimates landslide susceptibility, it will classify most areas as high susceptibility zones, or the percentage of areas under high susceptibility zone will be comparatively higher than other zones.

3.4. Results

3.4.1. Variable Importance of the Causal Factors

Variable importance shows which causal factors have the most predictive power in a random forest model (Chen et al., 2017). In our proposed MD-based sampling method (Fig 3.7), elevation (100.0) was the most important causal factor, followed by the distance from drainage network (75.7), distance from the fault lines (66.1), slope (61.6) and geology (50.1). Factors like profile curvature (0.0), NDVI (11.0) has the least importance in the model.

In the slope-based sampling (Fig 3.7), TWI (100.0) was the most important causal factor. Followed by the distance from the road network (86.8) and elevation (49.7). TWI is a slope product, and since in slope-based sampling, the slope was excluded from the model, TWI became the most important causal factor. Factors like aspect (0.0), SPI (9.3), and PR (17.4) were the least critical causal factors. SPI is another slope product; since TWI became an essential causal factor, another slope product had less importance in the model. If we compare the variable significance of MD and slope-based sampling, it is evident that the sampling method impacts deciding the causal factor's significance. For example, in MD-based sampling, elevation is the most important causal factor, but it was the third most important causal factor in the slope-based sampling method. In MD-based sampling, comparatively smaller areas than the slope-based sampling were used for absence-data sampling. But it was spread over a large area. On the other hand, in the slope-based sampling Kaptai lake, areas near Kaptai lake and areas with gentle slopes in the southwest were designated as a safe zone for absence-data sampling. Therefore, landslide locations were the same, but the outcome was different because of the difference in absence-data sampling.

3.4.2. Landslide Susceptibility Maps

Each landslide susceptibility map provides landslide probabilities from 0.0 to 1.0. We used a natural break method to classify the landslide probabilities into five susceptibility zones (Fig 3.8): very low, low, moderate, high, and very high.

In landslide susceptibility map produced using proposed MD-based sampling, valleys in the southeast areas (Fig 3.8) near the Rangamati lake were classified as either low or very low susceptibility zones. High and very high susceptibility zones spread around the surrounding areas of the landslides. There were high susceptibility zones in the north-west of the study areas. These areas contain Chittagong-Rangamati highway. It is because distance from the road network had higher variable importance in the random forest model in determining the landslide susceptibility. Elevation and slope were the other two important causal factors and that's why areas with comparatively higher elevation and steeper slope were classified either as high or very high susceptibility zones. At the same time distance from the fault lines was another causal factor which had comparatively higher variable importance in the model. In the study area there is a fault lines that stretched from the north-west to south-west and thus areas near to that fault was classified as either high or very high susceptibility zones.

On the other hand, for slope-based absence-data sampling Kaptai lake and areas near to the lake and some small patches in the south-east were classified as either very low or low susceptibility zones. From visual interpretation and comparison of slope and MD based methods it is evident that, in slope-based sampling method comparatively more areas were classified as either high or very high susceptibility zone than the MD-based sampling method. Some areas in the south east of the study area, were classified as low or moderate susceptibility zones but in slope-based absence-data sampling same areas were classified as either high or very high susceptibility zones. Moreover, in the study area, there is a fault line that stretches from the north west to south west. In this area elevation is also comparatively high than the other parts of the study area. In slope-based sampling these areas were classified as either high or very high susceptibility zone. But in MD-based sampling method in these areas there were patches of very high and high susceptibility zones. It did not classify the whole area as either high or very high susceptibility zones like the slope-based sampling method.

3.4.3. Performance of Landslide Susceptibility Maps

3.4.3.1. Success and Prediction Rates

In MD-based sampling the success and prediction rates (Fig 3.9.a) were 88.4% and 86.21% respectively. On the other hand, in slope-based sampling the success and prediction rates (Fig

3.9.b) were 89.51% and 88.71% respectively. For both success and prediction rates slope-based sampling outperformed MD-based sampling by 1.24% and 2.90%. Generally, the performance of a model is evaluated based on the prediction rate or how well it will predict the unknown landslides. From this perspective, the slope-based sampling is slightly better. However, the prediction and success rates are in good category of 80.0–90.0% in both the sampling methods, so that the difference in accuracy is not significant. From visual interpretation we can see that slope-based sampling classified comparatively more areas as high or very high susceptibility zones. These results suggest that the slightly high accuracy of the slope-based sampling is likely caused by the fact that this method classified more areas as high or very high susceptibility zones, an evidence of overestimation of landslide susceptibility of the slope-based sampling.

3.4.3.2. Statistical Index based Measures

For MD-based sampling TPR and TNR (Table 3.2) were 0.93 and 0.92, respectively. It means MD-based sampling method had same accuracy in differentiating the absence and presence data of landslides. TPR and TNR were 0.89 and 0.87 respectively for the validation data set. For the training dataset, the model attained strong agreement (Kappa >0.8), but for validation, the Kappa value was 0.75, which is moderate agreement.

In slope-based sampling for training dataset, TPR and TNR (Table 3.2) were 0.94 and 0.93, respectively. Like MD-based sampling, here the model showed similar performance in distinguishing absence and presence data. For validation dataset TPR and TNR were 0.90 and 0.93, respectively. Here, for unknown data, the model was 3.33% more accurate in detecting absence-data than detecting presence data. Kappa indices for the training and validation dataset were 0.84 and 0.83, respectively. Generally, the model that have the lowest difference in accuracy between the training and validation dataset is the best model.

3.4.3.3. Seed Cell Accuracy Index (SCAI)

SCAI assesses the consistency and desirability of the landslide susceptibility model. The SCAI value will decrease from very low to very high susceptibility zones (Arabameri et al., 2019). The model that has the lowest SCAI value for the very high susceptibility zone will be the most desirable (Abedini and Tulabi, 2018; Arabameri et al., 2019). It means the model will classify the least percentage of the area as a very high susceptibility zone, and most of the landslides will fall in this zone.

In the landslide susceptibility map produced by the MD-based sampling method, around 58.0% of the study area was classified as very low or low susceptibility zones, while around 35.0% of the study area was classified as high or very high susceptibility zones. The SCAI values decreased from 28.21 to 0.13 with the increase of the susceptibility from very low to very high. This indicates that the susceptibility map is consistent, and it classified a significant portion of the study area as very low or low susceptibility zones. A landslide susceptibility model overestimates landslide susceptibility when it cannot effectively differentiate high and low susceptibility zones. In particular, it misclassifies many low landslide susceptible areas to high susceptibility zones. This reduces the consistency and practical applicability of the model. The SCAI value of the MD-based model is low (0.13), indicating that it classified very few areas as high susceptibility zones where most of the landslides occurred. Therefore, the model is consistent.

In the landslide susceptibility map produced by the slope-based sampling, around 42.0% of the study area was classified as low or very low susceptibility zone, and around 46.0% of the study area was classified as high or very high susceptibility zones (Table 3.3). Compared to the MD-based sampling, the slope-based sampling classified almost two times more areas to high and very high susceptibility zones. Both slope and MD based sampling gave similar accuracy, but the landslide susceptibility map produced by the slope-based sampling classified almost half of the area as high and very high susceptibility zones. This indicates that the slope-based model may classify more low susceptible areas to high susceptible zones, a sign of overestimating landslide susceptibility. The SCAI value decreases with the increase of susceptibility. The SCAI value is 0.43 for the very high susceptibility area, which is 3 times of the SCAI value for the same zone produced by the MD-based sampling. Therefore, the landslide susceptibility map produced using the slope-based sampling is not as consistent and desirable as the map produced by the MD-based sampling of absence-data.

3.5. Discussion

We assessed the MD-based absence-data sampling method and compared it with the slope-based method for landslide susceptibility mapping. The MD values were compared with the chi-square distribution to determine the threshold for absence-data sampling. In MD-based sampling, the absence sampling space was spread over the entire study area. Since the whole dataset including landslide locations and 15 causal factors, the use of this sampling method does not bias towards any specific landslide location. Several other distance-based matrices like similarity index

have been used for absence-data sampling (Zhu et al. 2019), but it is still unclear how to choose the critical value to determine the safe zone. Our proposed method provides an objective means to determine the critical value based on the Chi-square distribution of the MD values of the landslide locations and a user-specified confidence level.

The slope-based threshold has been commonly used for absence-data sampling. However, unlike the MD-based sampling, it is impossible to determine the critical value for the slope-based sampling because the degree of freedom is zero. In our comparison study, the size of the sampling space based on the threshold of slope < 3^{o} was comparatively larger than the MD-based sampling, but the sampling space was more clustered in the Kaptia lake and its nearby area. Therefore, the absence-data based on the slope threshold were sampled only from these clustered areas, while absence-data was sampled from a variety of areas in the MD-based method. The slope-based sampling classified most areas as either very high or very low susceptibility zones. It also classified some landslide free zones as vulnerable zones, overestimating the landslide susceptibility of the area (Hong et al. 2018; Zhu et al. 2010).

The slope-based sampling has been widely used in landslide susceptibility mapping (Althuwaynee et al., 2014; Tsangaratos and Bernados, 2014). Some studies have also included slope in the model although it has already been used for absence-data sampling. The double counting of the slope factor likely produced a biased model. We recommend that when the slope is used in absence-data sampling, it should not be included in the model.

Success and prediction rates and statistical measures are generally used for accuracy assessment, and in most cases, the consistency and desirability of the map are ignored (Althuwaynee et al., 2014; Abedini and Tulabi, 2018; Zhu et al. 2019; Rabby et al. 2020). The threshold based on which the safe zone is determined generally depends on the accuracy (Zhu et al., 2019). However, the landslide susceptibility map may lose its consistency because the higher accuracy can be achieved by increasing the areas of high and very high susceptibility zones (Abedini and Tulabi, 2018). Therefore, both accuracy and consistency should be assessed. Our study showed that MD-based sampling provides a landslide susceptibility map with satisfactory accuracy and consistency. In contrast, the slope-based sampling may increase the accuracy, but damage the consistency because the model classified most areas as high susceptibility zones (Abedini and Tulabi, 2018).

For our proposed MD-based sampling method, the MD values can be compared with a probability distribution and a confidence level to determine the critical value. In contrast, the determination of the slope threshold is subjective. Our proposed method reduces the subjectivity in choosing the threshold. Our proposed method is more statistically robust and scientifically viable than the slope-based sampling.

3.6. Conclusions

This study proposed an objective MD-based absence-data sampling method for landslide susceptibility mapping. We compared our proposed method with a commonly used slope-based absence-data sampling in producing landslide susceptibility maps based on a random forest model. Our results indicate that the landslide susceptibility map produced using the MD-based method is satisfactory in accuracy and consistency. Our proposed approach is less subjective because the critical value was determined based on a Chi-square distribution and a user-specified significance level. On the other hand, the slope-based sampling is subjective and results in a biased model towards the slope. We recommend excluding the slope from the model if it is used in absence-data sampling. Although the slope-based method produces a better accuracy for landslide susceptibility map in terms of AUC and statistical indices, the SCAI values indicated this method overestimates landslide susceptibility. The slope-based absence-data sampling method depends on the researcher's judgment and is based on one landslide causal factor. In contrast, multiple factors are used in MD-based absence-data sampling to determine the sampling space. Therefore, our proposed MD-based sampling method is more objective ad statistically robust than the slope-based method. It can be used for landslide susceptibility mapping in other areas, especially where landslide inventory is not representative for the whole region.

References

Abedini, M. and Tulabi, S., 2018. Assessing LNRF, FR, and AHP models in landslide susceptibility mapping index: a comparative study of Nojian watershed in Lorestan province, Iran. *Environmental Earth Sciences*, *77*(11), pp.1-13.

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), pp.1707-1737.

Ahmed, B. and Dewan, A., 2017. Application of bivariate and multivariate statistical techniques in landslide susceptibility modeling in Chittagong city corporation, Bangladesh. Remote Sensing, 9(4), p.304.

Ahmed, B., Rahman, M., Islam, R., Sammonds, P., Zhou, C., Uddin, K. and Al-Hussaini, T., 2018. Developing a Dynamic Web-GIS Based Landslide Early Warning System for the Chittagong Metropolitan Area, Bangladesh. ISPRS International Journal of Geo-Information, 7(12), p.485.

Adnan, M.S.G., Rahman, M.S., Ahmed, N., Ahmed, B., Rabbi, M. and Rahman, R.M., 2020. Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sensing*, *12*(20), p.3347.Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble decision tree-based Chi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. Landslides, 11(6), pp.1063-1078.

Ali, S.A., Parvin, F., Vojteková, J., Costache, R., Linh, N.T.T., Pham, Q.B., Vojtek, M., Gigović, L., Ahmad, A. and Ghorbani, M.A., 2021. GIS-based landslide susceptibility modeling: A comparison between fuzzy multi-criteria and machine learning algorithms. *Geoscience Frontiers*, *12*(2), pp.857-876.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. Catena, 114, pp.21-36.

Althuwaynee, O.F., Pradhan, B. and Lee, S., 2016. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. International Journal of Remote Sensing, 37(5), pp.1190-1209.

Arabameri, A., Chen, W., Loche, M., Zhao, X., Li, Y., Lombardo, L., Cerda, A., Pradhan, B. and Bui, D.T., 2020. Comparison of machine learning models for gully erosion susceptibility mapping. *Geoscience Frontiers*, *11*(5), pp.1609-1620.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology, 65(1-2), pp.15-31

Banglapedia., 2015. Landslide, Banglapedia: National Encyclopaedia of Bangladesh. Asiatic Society of Bangladesh. http://en.banglapedia.org/index.php?title=Landslide. Accessed 10 March 2018

Bangladesh Bureau of Statistics (BBS)., 2011, Population census 2011, Rangamati: ministry of planning, Dhaka, Bangladesh.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Budimir, M.E.A., Atkinson, P.M. and Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides, 12(3), pp.419-436.

Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I. and Dick, O.B., 2012. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. Geomorphology, 171, pp.12-29.

Carrara, A., Crosta, G. and Frattini, P., 2008. Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology*, *94*(3-4), pp.353-378.

Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A.X., Pei, X. and Duan, Z., 2018. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Science of the total environment*, *626*, pp.1121-1135.

Chen, W., Zhang, S., Li, R. and Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Science of the total environment, 644, pp.1006-1018.

Chen, W., Shahabi, H., Shirzadi, A., Hong, H., Akgun, A., Tian, Y., Liu, J., Zhu, A.X. and Li, S., 2019. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment, 78(6), pp.4397-4419.

Cruden, D.M. and Varnes, D.J., 1996. Landslides: investigation and mitigation. Chapter 3-Landslide types and processes. Transportation research board special report, (247).

Davis, J.C. and Sampson, R.J., 2002. *Statistics and data analysis in geology* (Vol. 646). New York: Wiley.

Fabbri, A.G., Chung, C.J.F., Cendrero, A. and Remondo, J., 2003. Is prediction of future landslides possible with a GIS?. *Natural Hazards*, *30*(3), pp.487-503.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology*, *94*(3-4), pp.268-289.

Glade, T., 2001. Landslide hazard assessment and historical landslide data—An inseparable couple. In The use of historical data in natural hazard assessments (pp. 153-168). Springer, Dordrecht.

Guzzetti, F., 2002, October. Landslide hazard assessment and risk evaluation: Limits and prospectives. In *Proceedings of the 4th EGS Plinius Conference, Mallorca, Spain* (pp. 2-4).

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology, 81(1-2), pp.166-184.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), pp.42-66.

Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A.X., Chen, W. and Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, pp.399-413.

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*, pp.11-22.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Kanwal, S., Atif, S. and Shafiq, M., 2017. GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomatics, Natural Hazards and Risk*, *8*(2), pp.348-366.

Khan YA, Lateh H, Baten MA, Kamil AA (2012) Critical antecedent rainfall conditions for shallow landslides in Chittagong City of Bangladesh. Environ Earth Sci 67:97–106. https://www.researchgate.net/publication/257793592_Critical_antecedent_rainfall_conditions_for_shallow_landslides_in_Chittagong_City_of_Bangladesh. Accessed 29 Nov 2018

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R news*, *2*(3), pp.18-22.

Kissell, R. and Poserina, J., 2017. *Optimal sports math, statistics, and fantasy*. Academic Press.

Li, C., Yan, J., Wu, J., Lei, G., Wang, L. and Zhang, Y., 2019. Determination of the embedded length of stabilizing piles in colluvial landslides with upper hard and lower weak bedrock based on the deformation control principle. Bulletin of engineering geology and the environment, 78(2), pp.1189-1208.

Koch, G.S. and Link, R.F., 2002. *Statistical analysis of geological data*. Courier Corporation.

Masum, K.M. Rashid, M,M. Jashimuddin, M. Ara, S.J.G. (2011). Land use conflicts of Chittagong Hill tracts and indigenous hill people as victim in Bangladesh. J Gen Educ 1:62–71

https://www.researchgate.net/publication/235703063_Land_use_conflicts_of_Chittagong_Hill_Tracts_and_indigenous_hill_people_as_victim_in_Bangladesh. Accessed 29 Nov 2018

Nader, P., Honeine, P. and Beauseroy, P., 2014, September. Mahalanobis-based one-class classification. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T., 2014. Assessing the quality of landslide susceptibility maps–case study Lower Austria. Natural Hazards and Earth System Sciences, 14(1), pp.95-118.

Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., Le, T.M., Van Phong, T., Khoi, D.K., Shirzadi, A. and Bui, D.T., 2020. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto International*, *35*(12), pp.1267-1292.

Prabhakaran, S., 2020. *Mahalanobis Distance - Understanding the math with examples (python) - ML+*.[online] ML+. Available at: <https://www.machinelearningplus.com/statistics/mahalanobis-distance/> [Accessed 8 April 2020].

Prothom Alo (2017). Rangamati landslide death toll 118. Available at: https://en.prothomalo.com/bangladesh/news/151605/Rangamati-Landslide-death-toll-118 (Accessed: 22-Jan-2019)

Rabby, Y.W. and Li, Y., 2019. An integrated approach to map landslides in Chittagong Hilly Areas, Bangladesh, using Google Earth and field mapping. Landslides, 16(3), pp.633-645.

Rabby, Y.W., Hossain, M.B. and Abedin, J., 2020. Landslide Susceptibility Mapping in Three Upazilas of Rangamati Hill District Bangladesh: Application and Comparison of GIS-based Machine Learning Methods. *Geocarto International*, pp.1-24.

Rahman MS, Ahmed B, Huq FF, Rahman S, Al-Hussain TM (2016) Landslide inventory in an urban setting in the context of Chittagong Metropolitan, Area, Bangladesh. In the proceedings of 3rdInternational conference in Civil Engineering, 21–23 December 2016 CUET, Chittagong, Bangladesh.

Rahman, M.S., Ahmed, B. and Di, L., 2017. Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. Journal of Mountain Science, 14(10), pp.1919-1937.

Rasyid, A.R., Bhandary, N.P. and Yatabe, R., 2016. Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. Geoenvironmental Disasters, 3(1), p.19.

Reichenbach, P., Mondini, A.C. and Rossi, M., 2014. The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). Environmental management, 54(6), pp.1372-1384.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180, pp.60-91.

Regmi, A.D., Devkota, K.C., Yoshida, K., Pradhan, B., Pourghasemi, H.R., Kumamoto, T. and Akgun, A., 2014. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. Arabian Journal of Geosciences, 7(2), pp.725-742.

Roy, J. and Saha, S., 2019. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenvironmental Disasters*, *6*(1), p.11.

Sabokbar, H.F., Roodposhti, M.S. and Tazik, E., 2014. Landslide susceptibility mapping using geographically-weighted principal component analysis. Geomorphology, 226, pp.15-24.

Sabatakakis, N., Koukis, G., Vassiliades, E. and Lainas, S., 2013. Landslide susceptibility zonation in Greece. Natural Hazards, 65(1), pp.523-543.

Schicker, R. and Moon, V., 2012. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. Geomorphology, 161, pp.40-57.

Sdao, F., Lioi, D.S., Pascale, S., Caniani, D. and Mancini, I.M., 2013. Landslide susceptibility assessment by using a neuro-fuzzy model: a case study in the Rupestrian heritage rich area of

Steckler, M.S., Nooner, S.L., Akhter, S.H., Chowdhury, S.K., Bettadpur, S., Seeber, L. and Kogan, M.G., 2010. Modeling Earth deformation from monsoonal flooding in Bangladesh using hydrographic, GPS, and Gravity Recovery and Climate Experiment (GRACE) data. Journal of Geophysical Research: Solid Earth, 115(B8).

Sterlacchini, S., Ballabio, C., Blahut, J., Masetti, M. and Sorichetta, A., 2011. Spatial agreement of predicted patterns in landslide susceptibility maps. Geomorphology, 125(1), pp.51-61.

Thanh, L.N. and De Smedt, F., 2012. Application of an analytical hierarchical process approach for landslide susceptibility mapping in A Luoi district, Thua Thien Hue Province, Vietnam. Environmental Earth Sciences, 66(7), pp.1739-1752.

Süzen, M.L. and Doyuran, V., 2004. A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental geology*, *45*(5), pp.665-679.

Tsangaratos, P. and Benardos, A., 2014. Estimating landslide susceptibility through a artificial neural network classifier. Natural Hazards, 74(3), pp.1489-1516.

Vakhshoori, V. and Zare, M., 2016. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. Geomatics, Natural Hazards and Risk, 7(5), pp.1731-1752.

Wang, Q., Wang, Y., Niu, R. and Peng, L., 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, 9(9), p.938.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S. and Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides*, *13*(5), pp.839-856.

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*(4), pp.572-582.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences*, *35*(6), pp.1125-1138.

Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences, 61(4), pp.821-836.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T. and Hong, H., 2019. A similarity-based approach to sampling absence-data for landslide susceptibility mapping using data-driven methods. Catena, 183, p.104188.

Appendix

Table 3.1. Landslide Causal Factors used in this Study

Factor Name	Type	Data Source	Resolution	Reasons to choose
Elevation	Geophysical	Abedin et al. (202)	30m	Geomorphic, environmental, and anthropogenic processes depend on elevation (kanwal et al. 2016).
Slope	Geophysical	Abedin et al. (2020)	30m	With the increase of slope probability of slope failure increase (Chen et al. 2019).
Plan Curvature	Geophysical	Abedin et al. (2020)	30m	Affects the concentration of water over the surface after rainfall and thus can control the pore pressure of the soil (Ayalew and Yamagishi, 2005).
Profile Curvature	Geophysical	Abedin et al. (2020)	30m	Affects the concentration of water over the surface after rainfall and thus can control the pore pressure of the soil (Ayalew and Yamagishi, 2005).
Aspect	Geophysical	Abedin et al. (2020)	30m	Aspect involves how much sunlight an area will receive. Consequently, it has effects on several geomorphic processes, including erosion and evapotranspiration (Chen et al. 2018).
TWI	Hydrological	Rabby et al. (2020)	30m	Represents stream power of erosion (Kanwal et al. 2016).
SPI	Hydrological	Rabby et al. (2020)	30 m	Represents stream power of erosion (Kanwal et al. 2016).
Distance from Road Network	Anthropogenic	Rabby et al. (2020)	1000m	Road construction in the hilly areas alters the structure of the landscape, increasing the probability of landslides (Kanwal et al. 2016).
Distance from Drainage Network	Hydrological	Rabby et al. (2020)	1000m	The probability of landslides is generally high near the stream network (Chen et al., 2018).
Distance from the Fault lines	Geological	Rabby et al. (2020)	1000m	Fault lines show the zones of weakness where the probability of landslides is high (Rabby and Li, 2020).
Geology	Geological	Rabby et al. (2020)	1000m	Geological formations: Dihing and Dupi Tila are susceptible to landslides (Ahmed, 2015).
Rainfall	Hydrological	Abedin et al. (2020)	1000m	Excessive rainfall in a short time acts as a triggering factor (Althuwaynee et al., 2014).
Normalized Difference Vegetation Index (NDVI)	Environmental	Abedin et al. (2020)	30m	It shows the vegetation health and in a vegetated surface probability of landslide is low (Kanwal et al. 2016).
Land use/Land cover (2018)	Environmental	Abedin et al. (2020)	30m	One of the main driving factors of landslides in the study area (Abedin et al. 2020).
Land use/Land cover change	Environmental	Abedin et al. (2020)	30m	The rate of land use land cover change is high in the study area which creates conducive condition for landslides (Rabby et al. 2020).

Table 3.2. Statistical Measures of Random Forest Model for Different Thresholds of Mahalanobis Distance

Sampling Method	Dataset	TPR	TNR	Kappa
MD-based	Training	0.93	0.92	0.86
	Validation	0.89	0.87	0.75
Slope-based	Training	0.94	0.93	0.84
	Validation	0.90	0.93	0.83

Table 3.3. SCAI Values for each Susceptibility Zones of Mahalanobis Distance-based Landslide Susceptibility Mapping

Sampling Method	Susceptibility	Area (%)	Landslide (%)	SCAI Index
Mahalanobis Distance-based	Very Low	33.57	1.19	28.21
	Low	24.87	4.76	5.22
	Moderate	19.34	15.87	1.22
	High	15.10	21.83	0.69
	Very High	7.12	56.35	0.13
Slope-based	Very Low	32.55	0.0	-
	Low	9.41	2.38	3.95
	Moderate	8.63	3.97	2.17
	High	15.67	13.10	1.20
	Very High	33.75	80.56	0.42

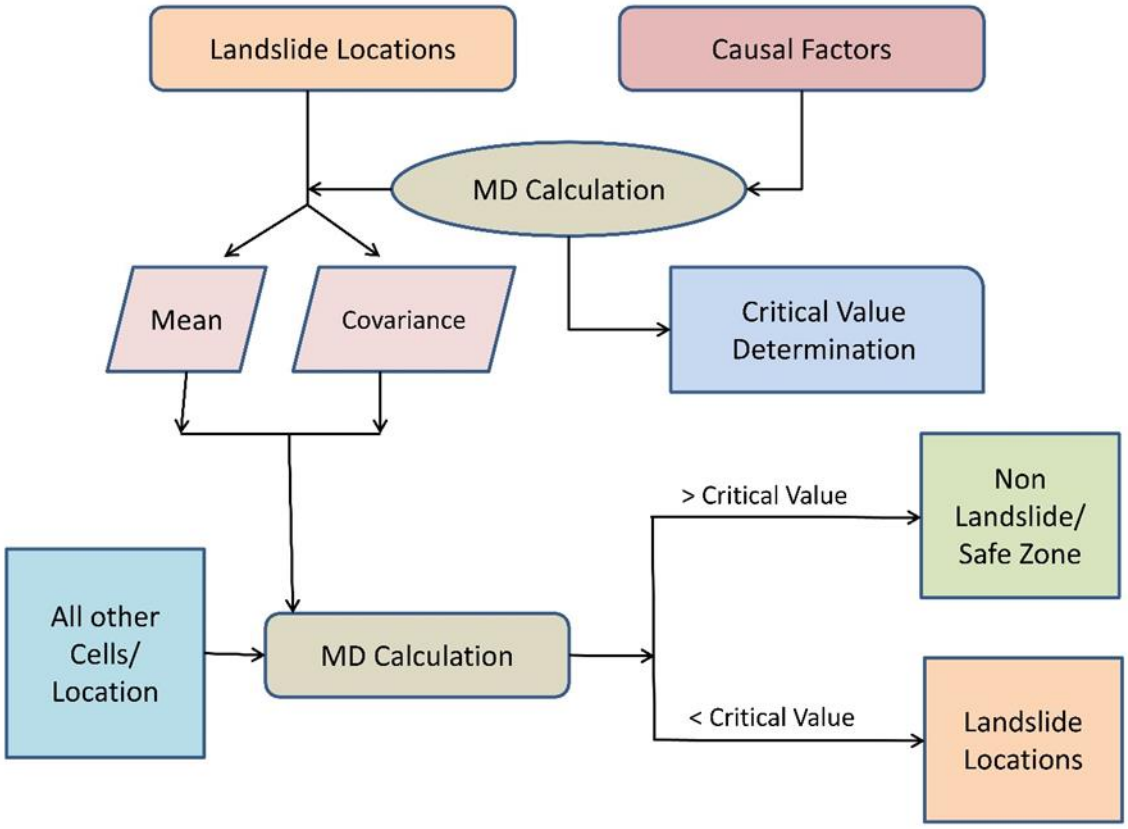

Fig 3.1. Flow Chart of the MD based Absence-data Sampling

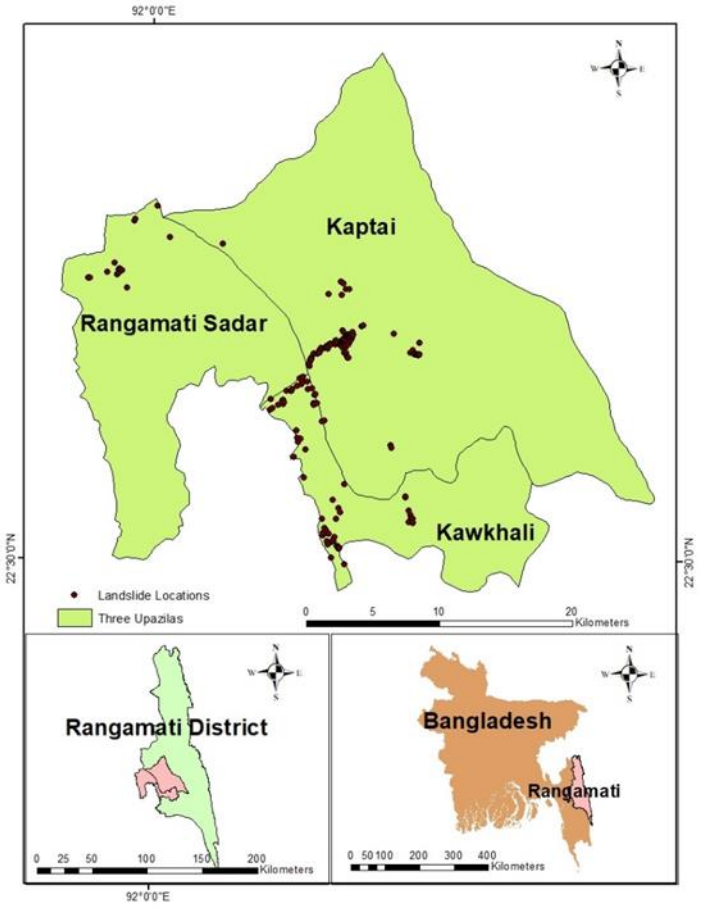

Figure 3.2. Study Area: Locations of Three Upazilas (Rangamati Sadar Kaptai and Kawkhali)

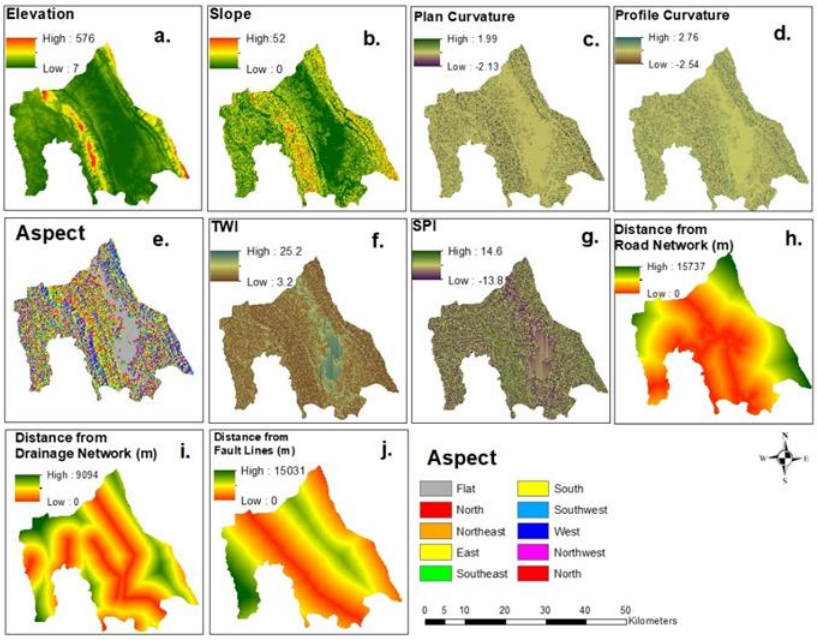

Fig 3.3. Landslide Causal Factors: a. Elevation; b. Slope; c. Plan Curvature; d. Profile Curvature; e. Aspect; f. TWI; g. SPI; h. Distance from the Road Network; i. Distance from the Drainage Network; j. Distance from Fault Lines

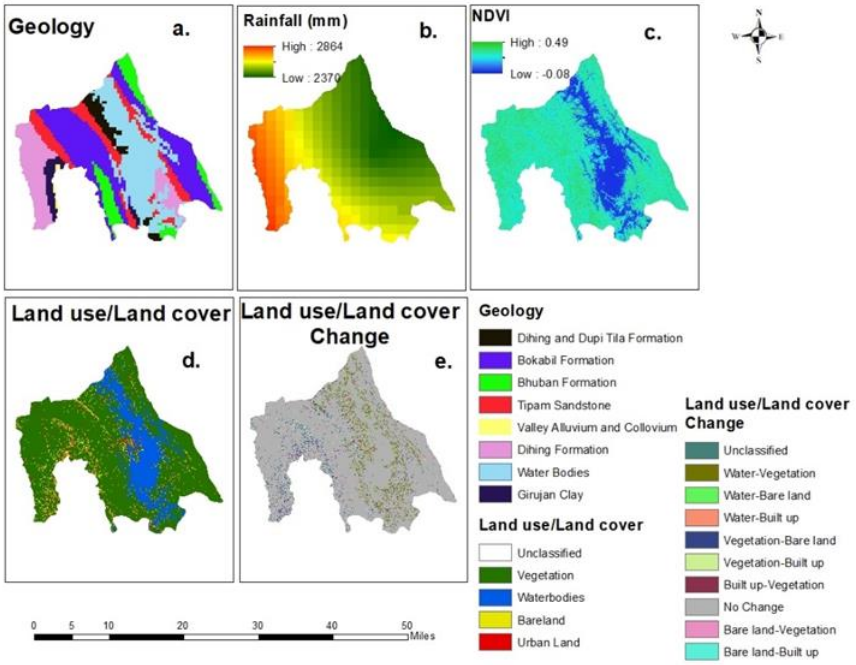

Fig 3.4. Landslide Causal Factors: a. Geology; b. Rainfall; c. NDVI; d. Land use/Land cover; e. Land use/Land cover Change

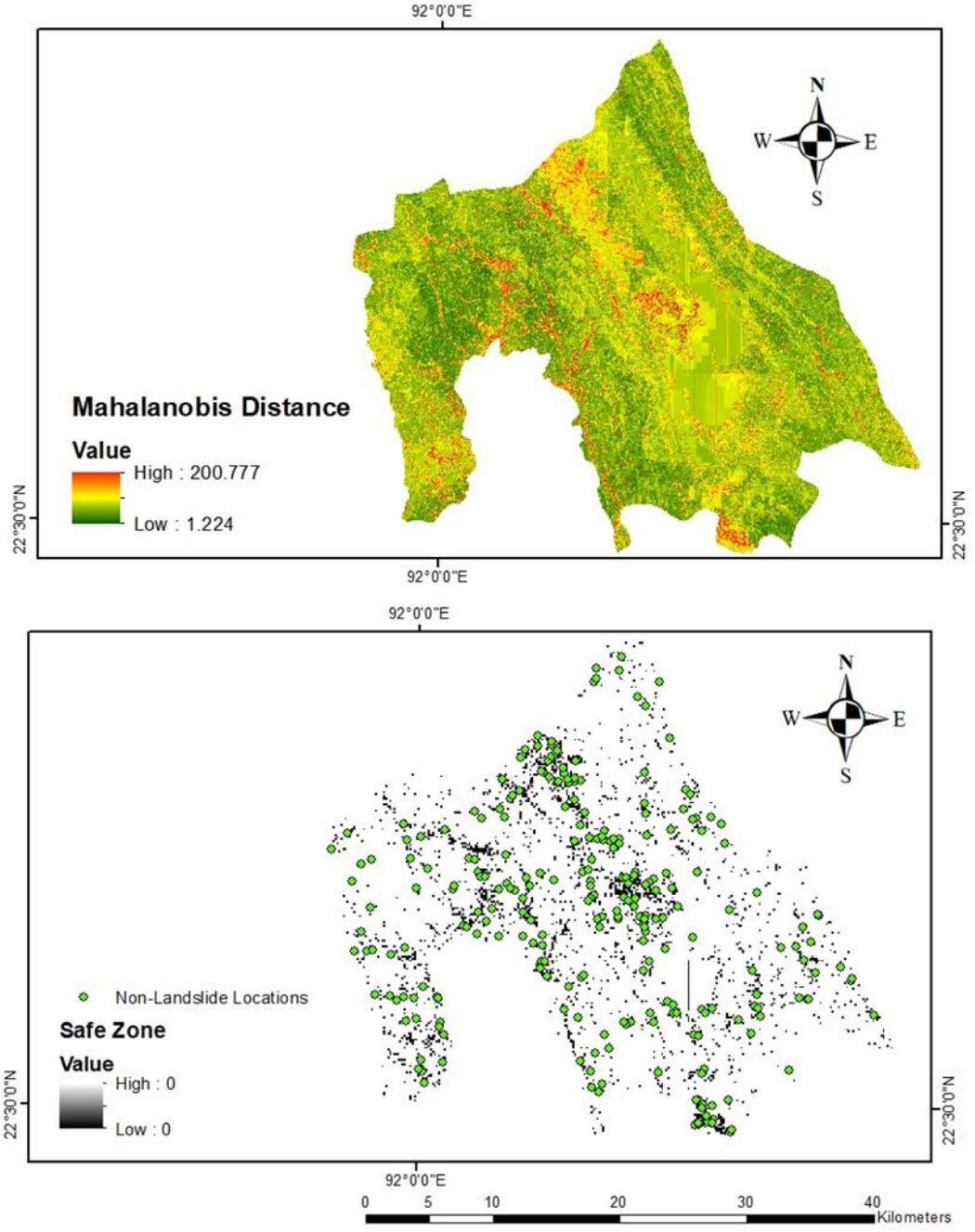

Fig 3.5. Spatial Distribution of Mahalanobis Distance (MD) and Sampling Space

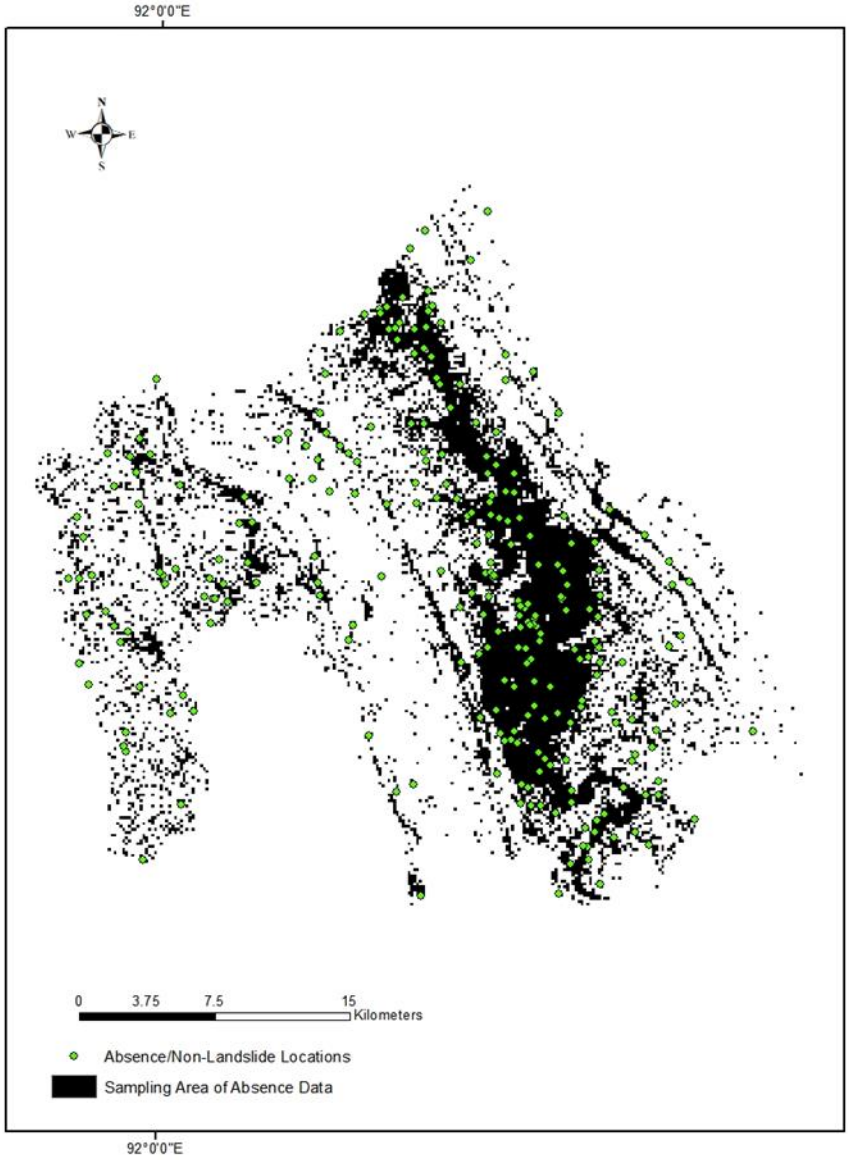

Fig 3.6. Absence-data Sampling Area based on Slope-Based Sampling

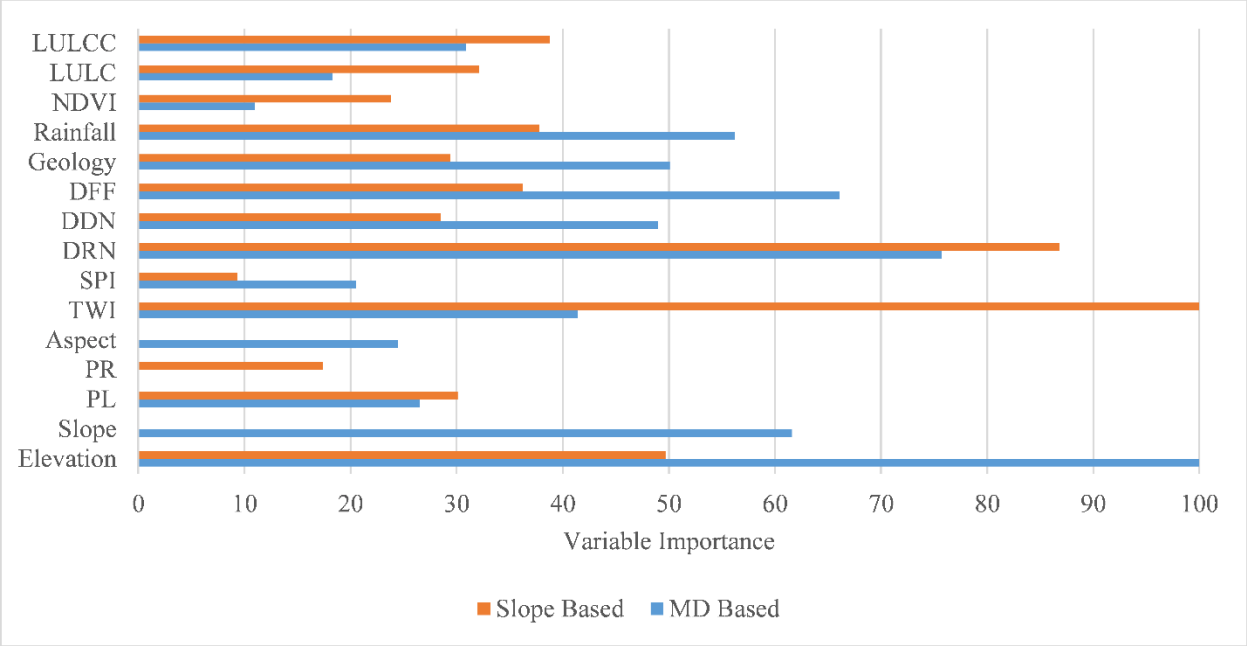

Fig 3.7. Variable Importance Plot of Random Forest Model Based on MD and Slope-based Absence Data Sampling

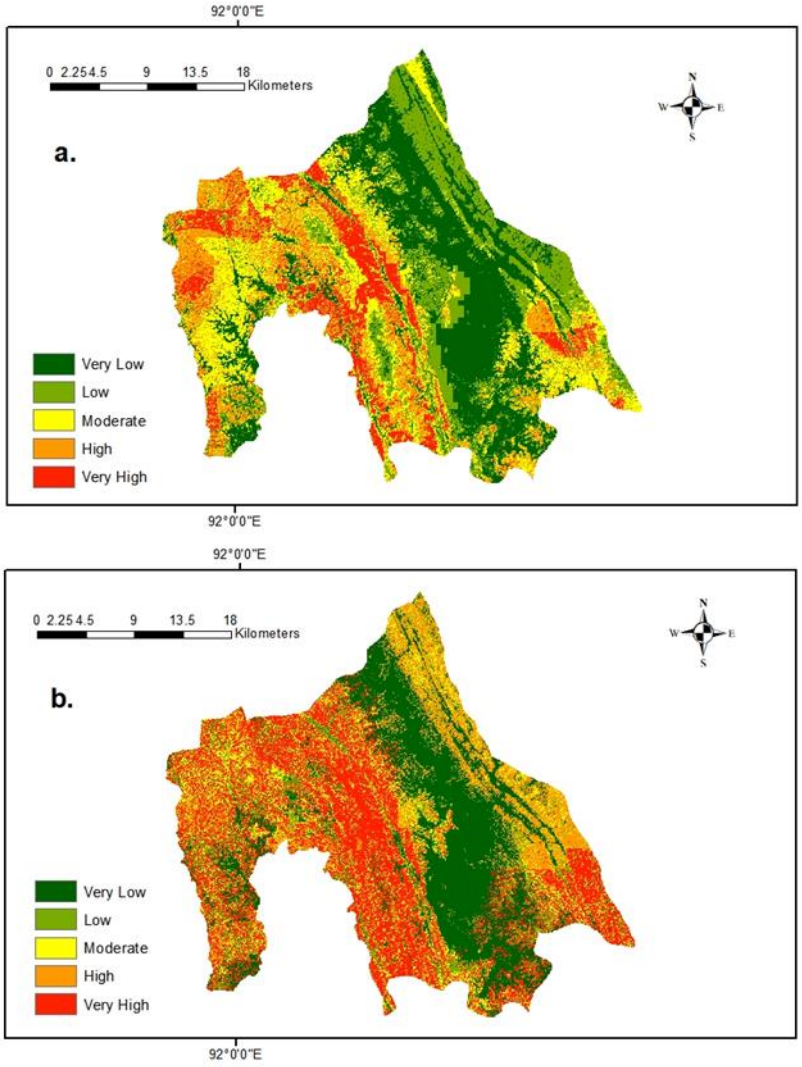

Fig 3.8. Landslide Susceptibility Maps based on the Random Forest Model using a. Mahalanobis Distance Based Absence-data Sampling; b. Slope-based Absence-data Sampling

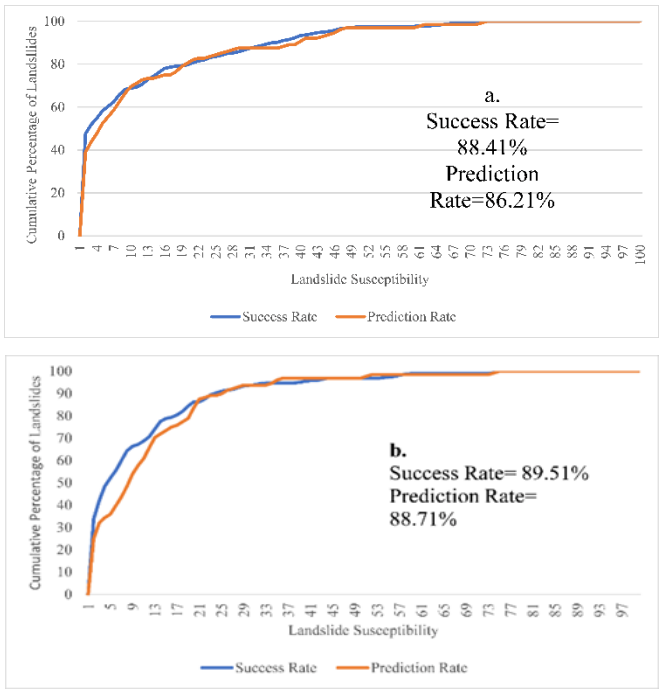

Fig 3.9. Success and Prediction Rate of Landslide Susceptibility Map based on a. Mahalanobis Distance Method b. Slope based Sampling

Chapter 4

Impact of Land use/ Land cover Change on Landslide Susceptibility in Rangamati Municipality of Rangamati District, Bangladesh

This chapter is a manuscript in preparation for *Science of Total Environment*. The use of "we" in this chapter refers to co-authors, Drs. Yingkui Li and me. As the first author, I conducted the fieldwork and data analysis and wrote the manuscript.

Abstract

Landslide susceptibility depends on various causal factors, such as geology, land use/land cover (LULC), slope, and elevation. Unlike other factors that are relatively stable over time, LULC is a dynamic factor associated with human activities. This study evaluates the impact of LULC change on landslide susceptibility in the Rangamati municipality of Rangamati district, Bangladesh, based on three LULC scenarios: the existing (2018) LULC; the proposed LULC (proposed in 2010, but has not been implemented yet); and the simulated LULC of 2028 using the artificial neural network (ANN) based cellular automata. The random forest model was used for landslide susceptibility mapping. The model showed good accuracies for all three LULC scenarios (Existing: 82.7%; Proposed: 81.4%; and 2028: 78.3%) and strong positive correlations (>0.8) between different landslide susceptibility maps. LULC is either the third or fourth most important factor in these scenarios, suggesting a moderate impact on landslide susceptibility. Future LULC changes likely increase the landslide susceptibility with up to 14.5% increases in the high susceptibility zone for both proposed and simulated LULC scenarios. These findings would help policymakers carry out proper urban planning and highlight the importance of considering landslide susceptibility in LULC planning.

4.1. Introduction

Landslides cause damage to infrastructure and casualties worldwide. As a representation of the spatial probability of landslides over an area (Reichenbach et al., 2018; Samia et al., 2018) landslide susceptibility mapping is critical to mitigating landslide disasters (Fell et al., 2008; Guzzetti et al. 2012; Segoni et al. 2018). Landslide susceptibility maps are produced using landslide inventory and causal factors (Guzzetti et al., 2005; Zhu et al., 2019; Dou et al. 2020). Landslide inventory shows the locations of landslides while landslide casual factors create suitable conditions for landslides (Ayalew and Yamagishi 2005; Guzzettie et al. 2012). Various statistical and machine learning models, including logistic regression, linear discriminate analysis, random forest, support vector machines, decision tree, extreme gradient boosting (XGBoost), frequency ratio, and certainty factor, have been used for landslide susceptibility mapping (Ayalew and Yamagishi, 2005; Nefeslioglu 2008; Bai et al. 2010; Regmi et al. 2014; Budimir et al. 2015; Zhang et al. 2018; Zhu et al. 2019; Rabby and Li, 2020). These models explore the relationships between landslide occurrences and causal factors to determine the spatial probability over the area (Althuwaynee et al., 2014; Dou et al., 2016; Reichenbach et al., 2018). Simple statistical models like logistic regression, frequency ratio, and certainty factor can produce easily understandable results with satisfactory accuracy (Ayalew and Yamagishi, 2005; Abedini and Tulabi, 2018). Advanced machine learning models like random forest and artificial neural network (ANN) usually produce much higher accuracy but less interpretability (James et al., 2013).

Landslide causal factors can be categorized into geological factors including lithology and distance from the fault lines, physiographic factors, such as slope, aspect, plan curvature, and profile curvature and environmental factors, like land use/land cover (LULC) and distance from the river (Kanwal et al. 2016). Geological and physiological factors are generally considered as static because they are relatively stable. In contrast, environmental factors, particularly LULC, are dynamic (Reichenbach et al., 2014). Different LULC types have different impacts on landslides. For example, vegetation usually stabilizes the slope because tree roots hold the soil together. Removing vegetation, either naturally or by anthropogenic activities, can create a conducive condition for landslides (Reichenbach et al., 2014). Similarly, infrastructural development like road construction alters slopes and causes landslides (Abedin et al. 2020).

Several studies have assessed the impact of LULC change, mainly deforestation, on landslide susceptibility (Genet et al., 2008; Reichenbach et al., 2014; Chen et al., 2019). Mainly associated

with agricultural activities, deforestation increases the weathering and erosion processes and ultimately increases the landslide susceptibility of an area (Mao et al., 2014). Chen et al. (2019) assessed the impact of LULC on landslide susceptibility based on decade wise LULC maps. Reichenbach et al. (2014) assessed the impact of different LULC scenarios on the landslide susceptibility in the Briga catchment of Messina, Italy. However, these studies simply used different LULC scenarios for the assessment without the consideration of the role of the LULC changing trend on the landslide susceptibility.

In recent days, machine learning methods have been used to simulate LULC change and the transitional potential of LULC types (Deng et al. 2009; Karimi et al., 2019; Hasan et al., 2020). The simulated LULC has been considered as the business as usual (BAU) scenario and this scenario reduces the subjectivity (Reichenbach et al. 2014; Mao et al. 2014; Chen et al. 2018). At the same time, LULC planning has been adopted to minimize the effects of natural hazards. Although it is generally assumed that planned LULC mitigates the impacts of natural hazards, few studies have evaluated the effects of a planned LULC on landslide susceptibility.

In this study, we assessed the impact of LULC change on the landslide susceptibility of Rangamati municipality, Bangladesh. Landslides occur mainly in the Chittagong Hilly Areas (CHA) (Ahmed 2015; Rabby and Li, 2019) in Bangladesh, especially in the three urban areas of Chittagong Metropolitan Area (CMA), Cox's Bazar, and Rangamati municipalities. These urban areas suffer from unplanned LULC change (Rahman et al., 2016; Rabby and Li; 2019; Abedin et al., 2020). Therefore, it is critical to assess future LULC changes on the landslide susceptibility in the Rangamati municipality. We evaluated the change of landslide susceptibility using the proposed LULC plan and simulated the LULC of 2028 (BAU). Specifically, this study helped answer the following research questions: a) what would be the landslide susceptibility scenario in BAU condition of LULC change and b) can planned LULC change prevent the increase of landslide susceptibility in the study area?

4.2. Materials and Methods

4.2.1. Study Area

Rangamati municipality (Fig 4.1) is the administrative headquarter of the Rangamati district. It covers approximately 64.8 km^2 of the area between 22º37'60 N and 91º2'0 E. The total population is around 150000, six times more than its carrying capacity. Population density is about

200 people/ km^2 (BBS, 2011). The elevation ranges from 0 to 195 m above sea level. The vegetation covers 75%, and water bodies cover 18% of the study area. Half of the study area is inhabitable, and that's why people clear forests and cut the hills to spread settlements and build new infrastructures (Prothom Alo, 2017). The maximum and minimum annual average temperatures are 36.5º C and 12.5 º C, respectively. The average annual rainfall is around 2673 mm, and (BBS, 2011).

Rangamati municipality is prone to landslides, and during June-July 2017, 73 people died due to landslides (Prothom Alo, 2017). The excessive monsoon rainfall triggered the landslides in a very short period (Abedin et al., 2020; Rabby et al., 2020). In the study area, population density has been doubled in the recent two decades because people migrated to this city (Prothom Alo, 2017). Due to the proximity to the Kaptai lake and natural scenic landscapes, tourism industries have started to grow. In recent years, plantation has become common in the city's western part (Abedin et al. 2020). Natural vegetation has been removed for plantation, increasing soil erosion (Prothom Alo, 2017). Unplanned LULC, infrastructural and tourism development and agriculture have increased the risk of landslides in this area.

4.2.2. Landslide Susceptibility Mapping

4.2.2.1. Landslide Inventory

We used 65 landslide locations (Fig 4.1) for susceptibility mapping. These landslides occurred during June-July 2017 and were mapped in the field. The same number of non-landslides (absence-data) were generated from the comparatively safer zones based on Mahalanobis distance-based absence-data sampling that we proposed in the previous chapter. These landslide and non-landslide locations were split into training (80%: 104) and validation (20%: 26) datasets.

4.2.2.2. Landslide Causal Factors

In this study, ten landslide causal factors: elevation, slope, aspect, topographic wetness index (TWI), stream power index (SPI), distance from the drainage network, plan curvature, profile curvature distance from fault lines, and LULC were used. We selected 30-m as the resolution for the landslide susceptibility map because most causal factors are with this resolution.

4.2.2.2.1 Relatively stable causal factors

Most causal factors we selected, except for the LULC, are relatively stable factors. The Advanced Spaceborne Thermal Emission Reflection Radiometer (ASTER) (30m×30m) DEM was

used to derive elevation, slope, aspect, plan curvature, and profile curvature (Fig 4.2.a-e). The slope is considered one of the most important causal factors because with the increase of slope, landslide probability increases (Kanwal et al. 2016). Other factors like pore pressure and water drainage also depend on the slope (Zhu et al., 2019). Aspect represents the direction of the slope. Profile curvature is defined as the parallel to the direction of the maximum slope. In contrast, plan curvature is perpendicular to the direction of maximum slope (ArcGIS 2020). These three factors may not directly affect the landslide susceptibility but, together with other factors, can create conducive conditions for landslides (Ahmed, 2015; Kanwal et al., 2016). Distance from the drainage network (Fig 2f) was derived from the drainage network downloaded from GeoDash, an open-access geospatial database provided by Bangladesh's government. Both the Topographic Wetness Index (TWI) (Fig 4.3.a) and Stream Power Index (SPI) are hydrological factors associated with the runoff potential (Kanwal et al. 2016), and they were also derived from the DEM (Fig 4.3.b). The map of fault lines from the Geological Survey of Bangladesh (GSB) was used to determine the distance from the fault lines (Fig 4.3.c). The closer distances to fault lines generally represent the weak locations with a high probability of landslides (Kanwal et al. 2016).

4.2.2.2.1 Land use/Landcover

Different from the stable factors described above, LULC is a dynamic factor. Abedin et al. (2020) found that LULC affects the landslide susceptibility of the study area. In particular, anthropogenic activities like plantation agriculture and urban infrastructure development cause rapid LULC change. In this paper, we examined the impact of three LULC scenarios on landslide susceptibility: (a) the existing (2018) LULC; (b) a proposed LULC; and (c) a simulated LULC of 2028.

Existing LULC of 2018

A Landsat 8 OLI image during the dry season (11/29/2018) was used to classify the LULC of 2018. The geometric and radiometric corrections were performed before the classification, and the image was reprojected to the Bangladesh Transverse Mercator System (BTM). We classified the image based on a modification of the Anderson Level-I scheme (Anderson, 1976). Before classification, all satellite data were studied using spectral and spatial profiles to ascertain the digital numbers (DNS) of different land use/cover categories. The classification scheme was established based on ancillary information of field survey, visual image interpretation, and local

knowledge of the study area. The classification of images was performed using a supervised maximum likelihood classification (MLC) algorithm. Based on the visual interpretation of the locations on Google Earth and the image itself, 60 polygons were digitized for each category. Using Rangamati district guide maps and Google Earth images, the land cover maps were validated. Four land-cover types, namely built-up, water bodies, vegetation, and bare land, were classified based on study area knowledge. Post-classification refinement was used to improve the classification accuracy (Dewan and Yamaguchi 2009). A 3*3 majority filter was also applied to reduce the salt-and-pepper effect to the classified maps (Lilesand and Kiefer 1999).

The classification accuracy was assessed using field data and the geographical features on land use maps, topographic maps from the survey of Bangladesh, and visual interpretation of very high spatial resolution data from Google Earth. The Landsat-derived classified images' total accuracy was 96%, with a corresponding kappa coefficient of 0.93. The user's and producer's accuracies of individual classes were ranging between 73-100% and 89-100%, respectively. The accuracy meets the standard of 85-90% for LULC mapping studies, as suggested by Anderson et al. (1976).

Proposed Land use/ Landcover

The second LULC scenario is a proposed LULC map by the town planning unit of Rangamati municipality under the "Urban Governance and Infrastructure Improvement Project." This proposed LULC has not been implemented yet. We aimed to assess whether the proposed LULC map can reduce the landslide susceptibility. The communication with urban planners of the municipality and the stakeholders indicates that the landslide susceptibility or the landslide risk was not considered when proposing the LULC map. However, all the rules of urban planning were used during the preparation of proposed LULC. For example, the new industrial and urban areas were proposed only in gentle slope areas.

We digitize this proposed land use map in ArcGIS. To be comparable with other LULC maps, we combined the LULC classes of the proposed map into four types: vegetation, water bodies, built up, and bare land.

Simulated Land use/ Land cover

The third LULC scenario is a simulated LULC in 2028. For LULC simulation, it is necessary to determine the factors that drive the LULC change of an area. These LULC classes are controlled

by different factors (Hasan et al., 2020). For Chittagong Hilly Areas (CHA), Hasan et al. (2020) used four categories of influencing factors: socio-economic, proximity to building infrastructure, climate, geophysical, and environmental factors. Since Rangamati municipality is situated in CHA. Table 4.1 shows the factors and their data sources used in the LULC simulation of 2028 in this study.

An artificial neural network (ANN) based cellular automata (ANN-CA) model was used to simulate and predict the LULC of 2028 based on the LULC change between 2008 and 2018. We assumed that the trend and dynamics of LULC changes would continue till 2028 (the BAU scenario).

Table 4.2 shows the transitional probability matrix of different LULC from 2008 to 2018. It is based on the percentage of LULC change from 2008 to 2018. The values in the matrix range from 0 to 1. The higher the value, the higher is the transitional probability of a land use type to convert into another type. The most active LULC type was vegetation and bare land since it had a higher probability of changing with (0.28) and built up (0.34). Waterbodies and built-up areas were the most stable type since Kaptai lake is a protected area in the study area. The probability of change of Kaptai lake is minimum. On the other built-up area will not convert back into vegetation or water bodies.

An open-source software package QGIS's MOLUSCE (Modules for Land-use Change Evaluation) plugin, was used to implement the ANN-CA model. This plugin measures the percentage of change area for each LULC of a study area. The transitional potential, calculated using the percentage of the change of LULC and its relationship with the influencing factors (Table 4.3), was used as the input in the cellular automata simulation of MOLUSCE plugins to predict future LULC (Saputra and Lee, 2019). ANN multilayer perception is used for transitional potential modeling, and the neural network had three layers: input hidden and output. In this study, eight exploratory variables were the input layers. Five hidden layers were used based on the 2n/3 approach where n=8. The learning rate was 0.001. The transitional probability provided by the ANN model was used in the cellular automata (CA) simulation for predicting the future LULC of 2028. In CA-based simulation, the composition and correlation of one cell with the surrounding cells are considered. CA-based simulation depends on the number of iterations, and the change of

LULC depends on the threshold value, ranging from 0 to 1. For a stable prediction, we set the threshold as 0.9 (Saputra and Lee, 2019).

4.2.2.3. Random Forest Model and Accuracy Assessment

The random forest model was used for landslide susceptibility mapping. Random forest is a widely used model in landslide susceptibility mapping since it shows better prediction capability (Chen et al., 2019; Zhu et al., 2019; Rabby et al., 2020). Random forest uses bootstrap aggregation and selects samples from the training dataset to develop a classification tree (James et al., 2013). Out of the bag samples or the unselected samples are used to determine the error and the importance of the model's factors (Zhu et al., 2019). The random forest model gives prediction by integrating individual classification trees (Chen et al., 2019; Pham et al., 2020). This model depends on two hyperparameters ntree or the number of trees and mtry or the number of nodes splits. For a stable model, ntree can be a large value and mtry=E/3 where E is the number of independent variables. “randomForest” package of R 3.8 was utilized to carry out the random forest modeling (Liaward and Weiner, 2002).

The area under the success and prediction rate curves were used for model validation. The training dataset was used to calculate the area under the success rate curve (AUC), while the validation dataset was used to calculate the area under the prediction rate. The AUCs of success and prediction rates range between 0.5 to 1.0 or 50% to 100% (Althuwaynee et al., 2014). 90-100% accuracy falls under the excellent category; 80-90% accuracy falls under the good category; 70-80% accuracy falls under the moderate category, and <70% falls under the poor category (Abedini and Tulabi, 2018). Model validation using the AUCs only assesses the accuracies of the models. It cannot show whether LULC maps in three different scenarios bring any change in the landslide susceptibility maps. We conducted the correlation analysis between the maps using the band collection statistics tool in ArcGIS to compare the three landslide susceptibility maps. If the use of different LULC maps brings substantial change to the appearance and susceptibility of the study area, then it is expected to have a low correlation between the maps.

4.3. Results

4.3.1. LULC Scenarios

4.3.1.1 Existing LULC of 2018

In 2018 (Fig 4.4.b), around 48.9% (Table 4.3) of the study area was designated as waterbodies. The percentage of vegetation covered around 36.5% of the study area. The percentage of built-up area and bare lands were 8.2% and 6.5%, respectively.

4.3.1.2 Proposed LULC

In the proposed LULC (Fig 4.4.d), around 38.4% were designated as either built-up areas or bare land. According to this proposed plan, some vegetation would be removed to develop industrial and commercial areas. Some areas in the southwest were designated as a fellow or bare land.

4.3.1.3 Simulated LULC in 2028

LULC of 2028 was simulated based on the trend of change of LULC from 2008 to 2018 and their association with the eight explanatory variables. From 2008 to 2018, vegetation decreased by 10.1% (Table 4.3), while bare land increased by 27.0% and built-up area increased by 88.9%. The increasing population and development of tourism industries are the reason behind the sharp increase of built-up area and decrease of vegetation. The simulated LULC pattern suggests that the build-up area would increase by 77.2%, and the bare land will increase by 54.8%. In contrast, vegetation would decrease by 4.9%, and the water bodies would reduce by 19.8% due to the conversion to the built-up or bare land.

4.3.2. Landslide Susceptibility Mapping

The variable importance plot (Fig 4.5) shows that, for the existing (2018) LULC, elevation (100.0) is the most important causal factor, followed by distance from the fault lines (65.5), distance from the drainage network (55.4), and LULC (55.3). In the proposed LULC scenario, elevation (100.0) was the most important causal factor, followed by distance from the fault lines (74.9) and distance from the drainage network (54.0). In this scenario, the importance of LULC (23.9) was not as high as the existing LULC scenario. For the simulated (BAU) LULC of 2028, elevation (100.0) was the most important causal factor, followed by distance from the drainage network (51.1), and distance from the fault lines (50.8).

Since elevation was the most important causal factor in the existing LULC scenario (Fig 4.6), areas with the higher elevation in the northwest and south-west regions were classified as either high or very high susceptibility zones. Simultaneously, the same areas near the built-up area were classified as either high or very high susceptibility zones. On the other hand, areas near the water bodies were classified as low susceptibility zones. For the proposed LULC map, the same areas were classified as either high or very high susceptibility zones. Moreover, high susceptibility zones spread around the classified high susceptibility zones by the existing LULC. In this scenario, the same areas near the water bodies and Kaptai lake were classified as moderate susceptibility zones. In the simulated LULC scenario (Fig 4.6), like the previous two models, the same areas were classified as high susceptibility zones and spread around the classified high susceptibility zones by the existing LULC. Like the proposed scenario, areas near the water bodies and the Kaptai lake were classified as moderate susceptibility zones because these areas were classified as the vegetation of built-up areas in the proposed and simulated LULC scenarios. In contrast, these areas were classified as water bodies in the existing LULC with low landslide probability.

In the existing LULC scenario (Table 4.4), 20.2% of the area was classified as high susceptibility zone. The high susceptibility zones were increased by 28.7% and 34.2% for the proposed and simulated LULC scenarios. Because only LULC was changed in three scenarios, the increase in the high susceptibility zones reflect the impact of LULC change, and its interaction with other factors.

The success (Table 4.5) (88.9%) and prediction rates (82.7%) were higher for the existing LULC than those of the other two LULC scenarios. The success rates are >80.0% for all scenarios. The prediction rates for the current and proposed LULC scenarios are relatively high (> 80%) than the rate for the simulated (2028) scenario (<80.0%). Table 4.6 shows positive correlations (>0.9) between the landslide susceptibility map produced for the existing LULC and the maps of proposed and simulated LULC scenarios. The variable importance of the random forest models shows similar ranking of the causal factors, resulting in high correlations between the landslide susceptibility maps.

4.4. Discussion

In this study, we assessed the impact of LULC on landslide susceptibility mapping in the Rangamati municipality based on three LULC scenarios. The landslide susceptibility map for the

existing LULC has the highest accuracy. The landslide locations used in this study occurred mainly during June-July 2017 and likely have a closer relationship with the existing LULC than the proposed LULC and simulated LULC of 2028.

In our study area, LULC is not the most significant factor for landslides. However, due to its dynamic nature, it can affect landslide susceptibility (Chen et al., 2019). Well-planned LULC can limit the increase of high susceptibility zones, and the business-as-usual scenario can exacerbate the condition (Reichenbach et al., 2014). Therefore, LULC change affects landslide susceptibility in the future.

In Rangamati municipality, the LULC changing rate is comparatively higher than the other parts of the Rangamati district. The random forest model showed that landslide susceptibility would increase for both proposed and simulated LULC scenarios, but the increase is lower in the proposed scenario. This suggests that the proposed LULC scenario is more sustainable than the BAU scenario. Although landslide susceptibility was not considered in the proposed LULC, the urban planning rules and regulations applied to the proposed LULC do mitigate the increase of landslide susceptibility. As mentioned before, in proposed LULC, the area under the built-up areas will increase, but new built-up areas will be proposed only in areas with gentle slopes. In contrast, BAU is dependent on the past trend of the LULC change. If the LULC changing trend continues, LULC will likely elevate the landslide susceptibility much higher. The changes in the built-up and bare lands are similar for the proposed and BAU scenarios. In BAU, the analysis was conducted at the pixel level, leading to more sporadic changes. In contrast, the proposed LULC was vector-based with large and continuous areas designated for a single LULC type. For example, the southwest portion of the study area includes four LULC types in the BAU scenario, but only two LULC types in the proposed LULC.

In future scenarios, BAU will increase the percentage of areas under high susceptibility zones. It is also evident that new high susceptibility zones will spread around the already classified susceptibility zones by the existing landslide susceptibility map based on LULC of 2018. It indicates that high susceptibility zones will not shift to an entirely new place; instead, it will spread around the previous locations.

Previous studies have found that LULC plays an essential role in determining the landslide susceptibility in this area (Rahman et al., 2016; Abedin et al. 2020). Our study confirmed previous

studies and suggested that the impact of LULC will increase in the future scenarios. The quality of the landslide susceptibility map depends on the quality and accuracy of landslide inventory and the causal factors (Kanwal et al., 2016; Guzzetti et al., 2012). In this study, 65 landslide locations were used for training and validating the models. These landslides occurred in 2017 and most of the landslides were near the settlement and other infrastructures like road networks. Because these landslides caused infrastructural damages and causalities, they were reported in newspapers and governmental reports. The resolution of the available satellite images was not high enough and some landslides may not be detected. Thus, the landslides were clustered in specific areas and may be not representative of the whole area. To reduce the biases, we excluded the factors like distance from the road network from the model. Due to the lack of high-resolution rainfall data, rainfall was not included as a causal factor in landslide susceptibility mapping. Geology is another critical factor that determines the susceptibility of an area. However, Rangamati municipality does not have a detailed geological map. Therefore, geology was excluded from the model.

This study only assessed the impact of LULC change on landslide susceptibility with the assumption that all other factors are unchanged. We acknowledge that other dynamic factors may also affect landslide susceptibility scenario. In particular, climate change may affect the pattern of rainfall, leading to the changes in landslide susceptibility. In our study area, landslides are mainly triggered by the intensive rainfall and climate change may result in more or less intensive rainfall events in the future. More studies are needed to assess the impact of climate change on landslide susceptibility.

4.5. Conclusions

In landslide susceptibility mapping, geomorphic and physiographic factors like slope, aspect, plan curvature, profile curvature, and geology are static. On the other hand, LULC a dynamic factor and is related to human activities. We assessed the impact of LULC change on landslide susceptibility based on three scenarios: existing, proposed, and simulated LULC patterns. The random forest model showed that due to LULC change, landslide susceptibility would increase, and thus the percentage of high susceptibility zone would also increase. All models showed satisfactory accuracy (>80.0%) for both success and prediction rates. The landslide susceptibility maps produced using three LULC scenarios had a very strong correlation. Future landslide susceptibility would keep changing with high susceptibility zones spreading around the existing high susceptibility zones mainly in the urban areas and areas with high elevation in the north and

southeast of the study area. A proper LULC management plan should be implemented to minimize the increase of high susceptibility zones. This study highlighted that high susceptibility zone likely spreads around existing high susceptibility zones. Proper LULC management policy is necessary to mitigate the increase of the high susceptibility zones.

In this study, we did not use causal factors, such as geology, rainfall, and soil characteristics, in landslide susceptibility mapping due to data unavailability. We also did not consider climate change in the assessment. Therefore, the produced landslide susceptibility maps may have some bias and uncertainties. Future work is necessary to include more factors in the assessment and assess the impact of climate change on landslide susceptibility.

Our results suggest that the proposed LULC scenario may have relatively lower increase in landslide susceptibility compared to the BAU scenario. However, it is unclear if the proposed LULC minimize the landslide susceptibility. It is important to explore different LULC scenarios to minimize landslide susceptibility in LULC planning and management.

References

Abedini, M. and Tulabi, S., 2018. Assessing LNRF, FR, and AHP models in landslide susceptibility mapping index: a comparative study of Nojian watershed in Lorestan province, Iran. *Environmental Earth Sciences, 77*(11), pp.1-13.

Abedin, J., Rabby, Y.W., Hasan, I. and Akter, H., 2020. An investigation of the characteristics, causes, and consequences of June 13, 2017, landslides in Rangamati District Bangladesh. *Geoenvironmental Disasters*, *7*(1), pp.1-19.

Ahmed, B., 2015. Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides, 12*(6), pp.1077-1095.

Anderson, J.R., 1976. *A land use and land cover classification system for use with remote sensor data* (Vol. 964). US Government Printing Office.

Althuwaynee, O.F., Pradhan, B., Park, H.J. and Lee, J.H., 2014. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena, 114*, pp.21-36.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, *65*(1-2), pp.15-31.

Bai, X., McAllister, R.R., Beaty, R.M. and Taylor, B., 2010. Urban policy and governance in a global environment: complex systems, scale mismatches and public participation. *Current opinion in environmental sustainability*, *2*(3), pp.129-135.

Bangladesh Bureau of Statistics (BBS)., 2011.Population census 2011, Rangamati: ministry of planning, Dhaka, Bangladesh.

Budimir, M.E.A., Atkinson, P.M. and Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. *Landslides, 12*(3), pp.419-436.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D.T., Duan, Z., Li, S. and Zhu, A.X., 2018. GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *Catena*, *164*, pp.135-149.

Chen, L., Guo, Z., Yin, K., Shrestha, D.P. and Jin, S., 2019. The influence of land use and land cover change on landslide susceptibility: a case study in Zhushan Town, Xuan'en County (Hubei, China). *Natural hazards and earth system sciences*, *19*(10), pp.2207-2228.

Deng, J.S., Wang, K., Hong, Y. and Qi, J.G., 2009. Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. *Landscape and urban planning*, *92*(3-4), pp.187-198.

Dewan, A.M. and Yamaguchi, Y., 2009. Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Applied geography*, *29*(3), pp.390-401.

Dou, J., Yunus, A.P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B.T. and Yamagishi, H., 2020. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Science of the total environment*, *720*, p.137320.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. and Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering geology*, *102*(3-4), pp.99-111.

Genet, M., Kokutse, N., Stokes, A., Fourcaud, T., Cai, X., Ji, J. and Mickovski, S., 2008. Root reinforcement in plantations of Cryptomeria japonica D. Don: effect of tree age and stand structure on slope stability. *Forest ecology and Management*, *256*(8), pp.1517-1526.

Guzzetti, F., Salvati, P. and Stark, C.P., 2005. Evaluation of risk to the population posed by natural hazards in Italy. *Landslide risk management, edited by: Hungr, O., Fell, R., Couture, R., and Eberhardt, E., Taylor & Francis Group, London*, pp.381-389.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, *112*(1-2), pp.42-66.

Hasan, S.S., Sarmin, N.S. and Miah, M.G., 2020. Assessment of scenario-based land use changes in the Chittagong Hill Tracts of Bangladesh. *Environmental Development*, *34*, p.100463.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Karimi, F., Sultana, S., Babakan, A.S. and Suthaharan, S., 2019. Urban expansion modeling using an enhanced decision tree algorithm. *GeoInformatica*, pp.1-17.

Kanwal, S., Atif, S. and Shafiq, M., 2017. GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomatics, Natural Hazards and Risk*, *8*(2), pp.348-366.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R news*, *2*(3), pp.18-22.

Lillesand, T., Kiefer, R.W. and Chipman, J., 2015. *Remote sensing and image interpretation*. John Wiley & Sons.Mao, D. and Cherkauer, K.A., 2009. Impacts of land-use change on hydrologic responses in the Great Lakes region. *Journal of Hydrology*, *374*(1-2), pp.71-82.

Nefeslioglu, H.A., Gokceoglu, C. and Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology*, *97*(3-4), pp.171-191.

Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., Le, T.M., Van Phong, T., Khoi, D.K., Shirzadi, A. and Bui, D.T., 2020. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto International*, *35*(12), pp.1267-1292.

Prothom Alo.; 2017. Rangamati landslide death toll 118. https://en.prothomalo.com/bangladesh/news/151605/Rangamati-Landslide-death-toll-118 (Accessed on January 22 2019)

Rabby, Y.W. and Li, Y., 2019. An integrated approach to map landslides in Chittagong Hilly Areas, Bangladesh, using Google Earth and field mapping. *Landslides*, *16*(3), pp.633-645.

Rabby, Y.W. and Li, Y., 2020. Landslide inventory (2001–2017) of Chittagong hilly areas, Bangladesh. *Data*, *5*(1), p.4.

Rabby, Y.W., Hossain, M.B. and Abedin, J., 2020. Landslide Susceptibility Mapping in Three Upazilas of Rangamati Hill District Bangladesh: Application and Comparison of GIS-based Machine Learning Methods. *Geocarto International*, pp.1-24.

Rahman, M.S., Ahmed, B. and Di, L., 2017. Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science*, *14*(10), pp.1919-1937.

Reichenbach, P., Mondini, A.C. and Rossi, M., 2014. The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). *Environmental management*, *54*(6), pp.1372-1384.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M. and Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, pp.60-91.

Regmi, N.R., Giardino, J.R., McDonald, E.V. and Vitek, J.D., 2014. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. *Landslides*, *11*(2), pp.247-262.

Saputra, M.H. and Lee, H.S., 2019. Prediction of land use and land cover changes for north sumatra, indonesia, using an artificial-neural-network-based cellular automaton. *Sustainability*, *11*(11), p.3024.

Samia, J., Temme, A., Bregt, A.K., Wallinga, J., Stuiver, J., Guzzetti, F., Ardizzone, F. and Rossi, M., 2018. Implementing landslide path dependency in landslide susceptibility modelling. *Landslides*, *15*(11), pp.2129-2144.

Segoni, S., Pappafico, G., Luti, T. and Catani, F., 2020. Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. *Landslides*, pp.1-11.

Zhang, S., Li, R., Wang, F. and Iio, A., 2019. Characteristics of landslides triggered by the 2018 Hokkaido Eastern Iburi earthquake, Northern Japan. *Landslides*, *16*(9), pp.1691-1708.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T. and Hong, H., 2019. A similarity-based approach to sampling absence-data for landslide susceptibility mapping using data-driven methods. *Catena*, *183*, p.104188.

Appendix

Table 4.1: Influencing Factors of LULC in Rangamati Municipality

Factor Type	Influencing Factor	Data Source
Socioeconomic Factors	Population Density	LandScan Project
Proximity to Build Infrastructures	Distance from the Road Network	GeoDash
	Distance from Urban Areas	Landsat 8
Climatic Variables	Rainfall	Bangladesh Meteorological Department (BMD)
	Elevation	ASTER (30m)
	Slope	ASTER (30 m)
	NDVI	Abedin et al. 2020
	Distance from Drainage Network	GeoDash

Table 4.2: Transitional Probability Matrix of Different Land use/Land covers in the Rangamati Municipality from 2008 to 2018.

LULC types	Waterbodies	Vegetation	Bare land	Built up
Waterbodies	0.90	0.09	0.0	0.0
Vegetation	0.04	0.36	0.28	0.34
Bare land	0.0	0.01	0.93	0.06
Built up	0.0	0.08	0.04	0.88

Table 4.3: Percentage of LULC Change in Different LULC Scenarios

Scenario	Year	Waterbodies (%)	Vegetation (%)	Built-up (%)	Bare land (%)
	2008	50.2	40.2	4.3	5.1
Base Year	2018	48.9	36.5	8.2	6.5
Business as Usual (BAU)	2028	40.8	30.7	14.5	10.6
Proposed		46.7	19.2	14.9	19.2

Table 4.4: Percentage of Area Under Different Susceptibility Zones Random Forest Model.

Model	**Land use**	**Susceptibility**	**Area (%)**
Random Forest	Existing	Low	63.6
		Moderate	16.2
		High	20.2
	Proposed	Low	59.0
		Moderate	15.0
		High	26.0
	2028 (Simulated)	Low	53.0
		Moderate	19.9
		High	27.1
		Moderate	6.6
		High	28.9

Table 4.5: Success and Prediction Rates of Random Forest Models.

Model	**Land use Data**	**Success Rate**	**Prediction Rate**
Random Forest	Existing	88.9	82.7
	Proposed	87.0	81.4
	2028	84.1	78.3

Table 4.6: Overall Correlation Between the Susceptibility Maps produced using Random Forest Model and Three Land use/Land Cover Scenarios

Model	**Land use**	**Existing**	**Proposed**	**2028**
Random Forest	Land use	Existing	Proposed	2028
	Existing	1.00	0.92	0.93
	Proposed	0.92	1.00	0.88
	2028	0.93	0.88	1.00

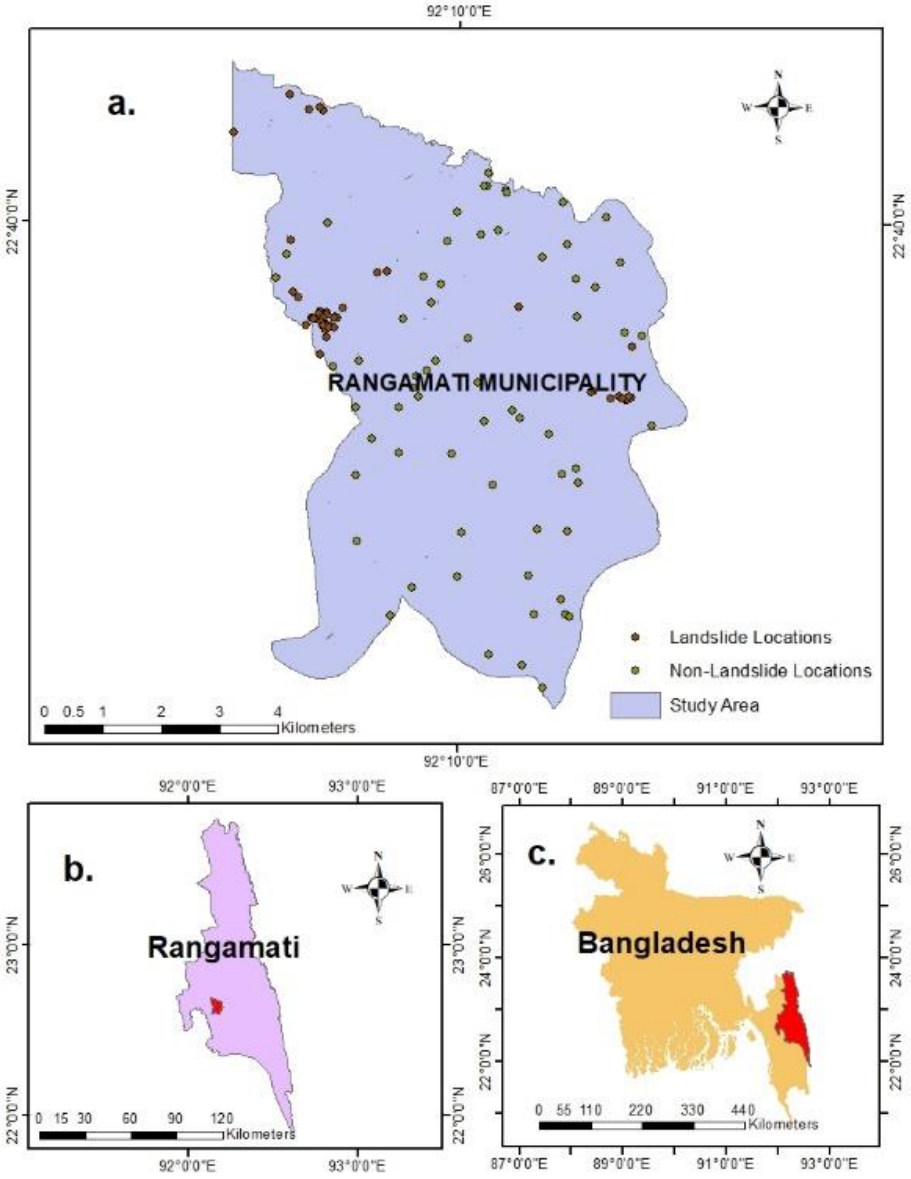

Fig 4.1. Location of Rangamati Municipality in Rangamati District, Bangladesh

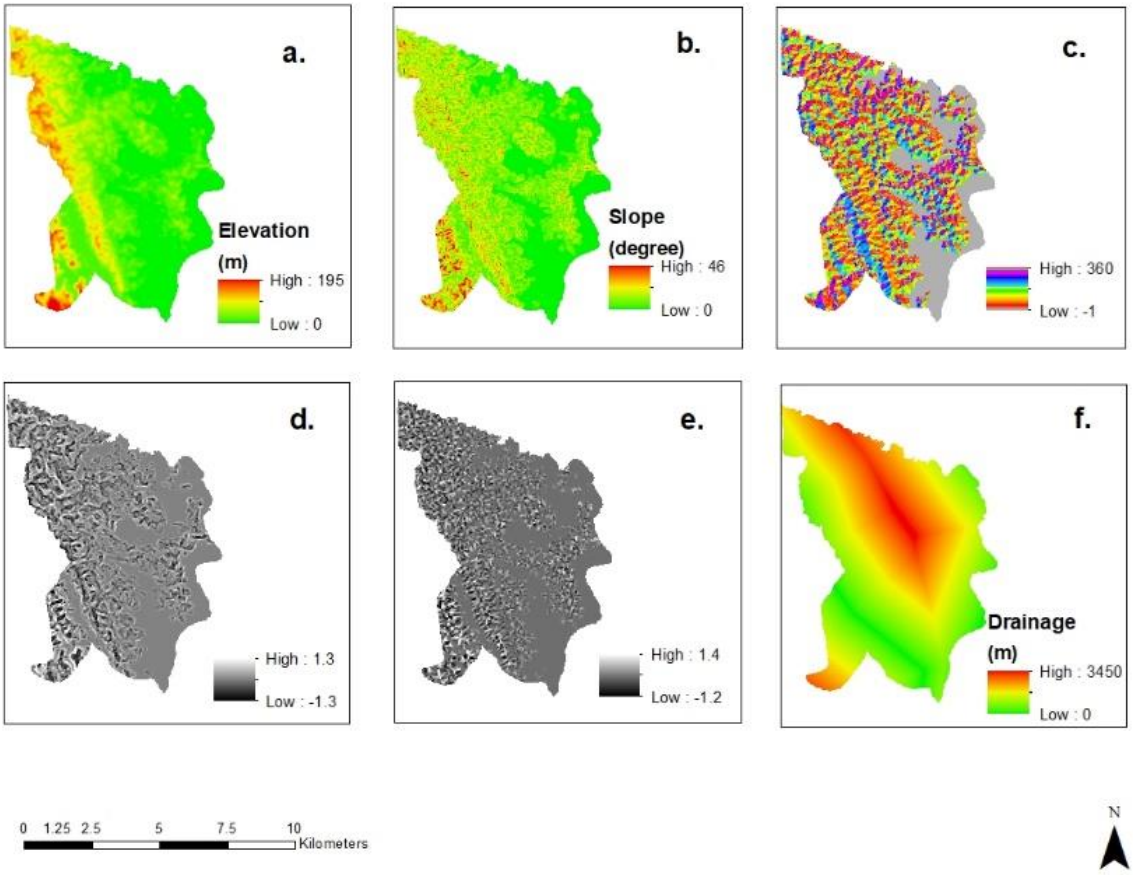

Fig 4.2. Landslide Causal Factors: a. Elevation; b. Slope; c. Aspect; d. Plan Curvature; e. Profile Curvature; f. Distance from the Drainage Network

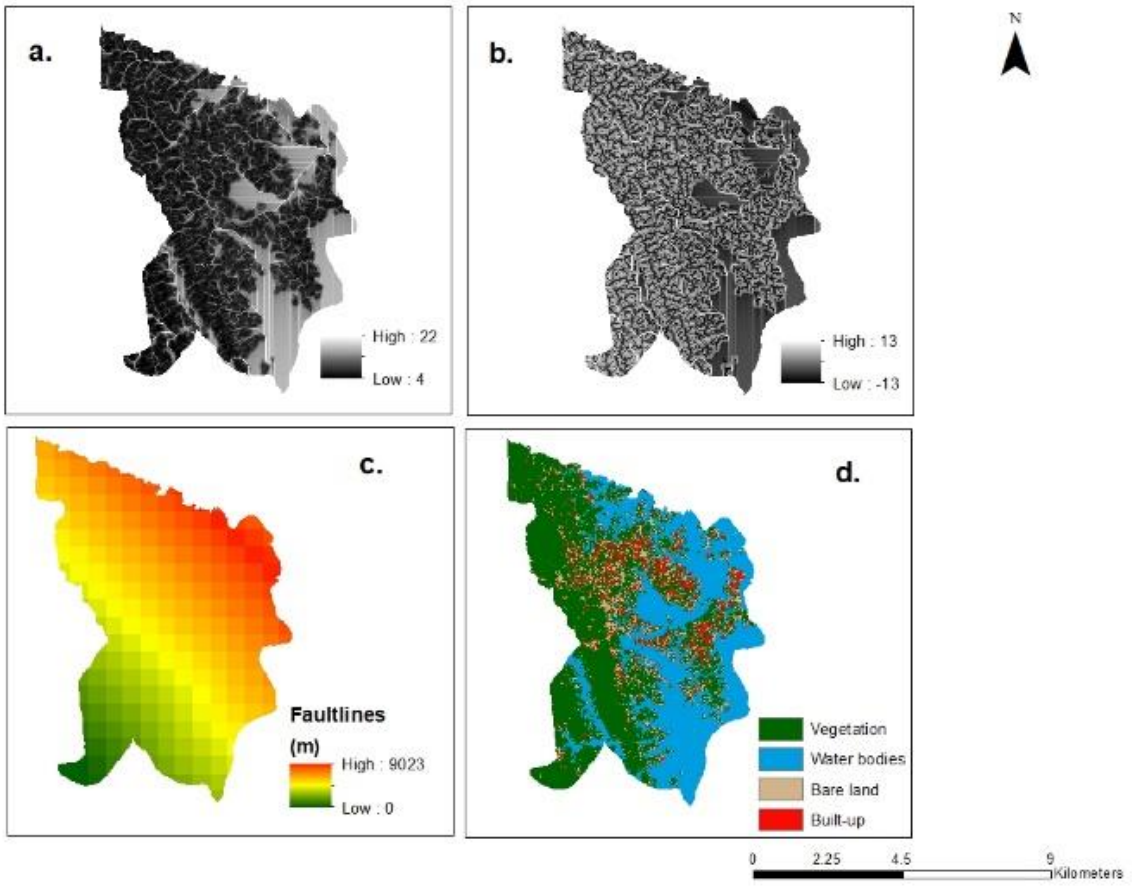

Fig 4.3. Landslide Causal Factors: a. TWI; b. SPI; d. Distance from the Faultline; d. Land use/Land cover

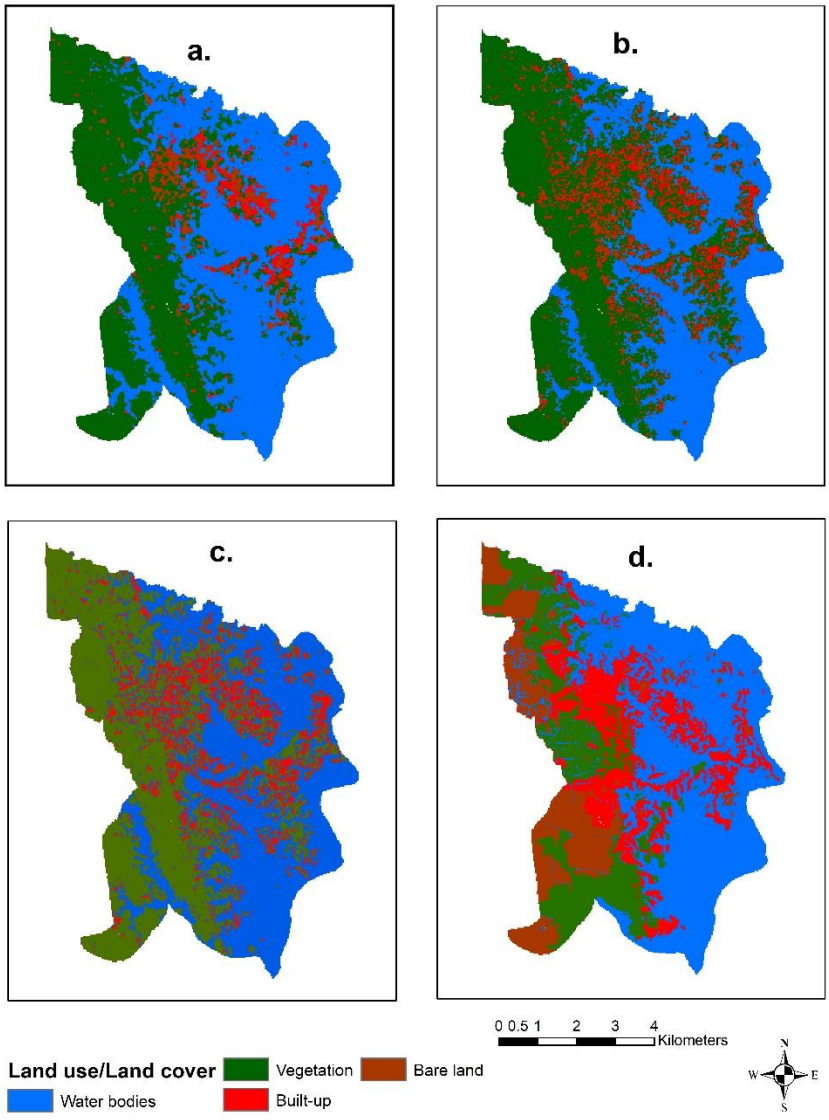

Fig 4.4: Land use/ Land cover Maps: a. Land use/ Land cover of 2008; b. LULC of 2018; c. Simulated LULC (2028) d. Proposed Land use/Land cover

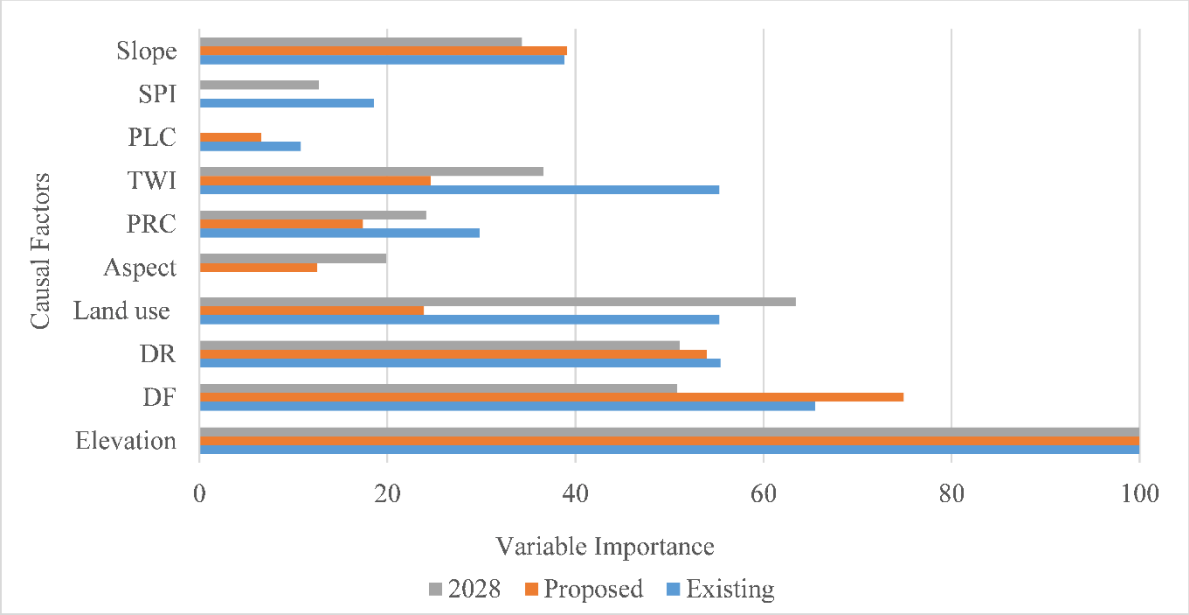

Fig 4.5: Variable Importance Plot for Random Forest Models based on Three LULC Scenarios

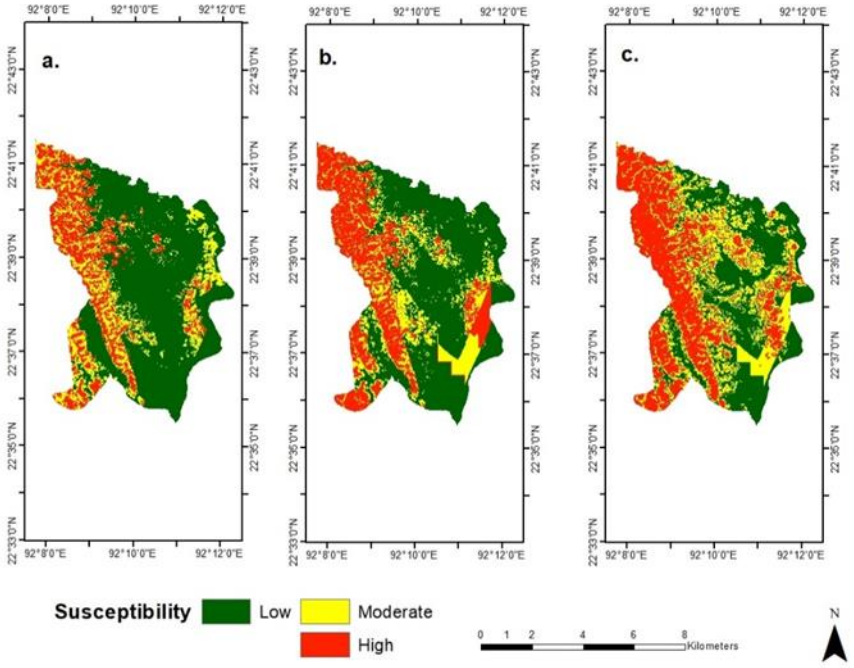

Fig 4.6: Landslide Susceptibility Maps Based on Random Forest: a. Existing Land use/Land cover; b. Proposed Land use/Land cover; c. Simulated Land use/ land cover of 2028

Chapter 5

Summary and Future Work

5.1. Summary and Major Findings

This dissertation research presented an integrated approach for landslide mapping using Google Earth and field mapping for CHA, Bangladesh. I also developed an MD-based absence-data sampling for landslide susceptibility mapping and applied this method in the landslide susceptibility mapping of three Upazilas of Rangamati district, Bangladesh. Finally, the impact of LULC change on landslide susceptibility mapping was evaluated in the Rangamati municipality of Rangamati district, Bangladesh.

CHA is prone to landslides, but no landslide inventory is available for the whole region (Ahmed, 2015; Rahman et al., 2017; Ahmed et al., 2020; Rabby and Li, 2020). This study produced a useful landslide inventory of CHA, which can be used for landslide susceptibility mapping of the whole area. In CHA, landslide inventories are only available in cities and towns, such as Chittagong Metropolitan and Cox's Bazar, based on field mapping (Ahmed 2015; Rahman et al. 2016). In this research, I prepared a landslide inventory for the whole area using an integrated method. I identified 230 landslides in Google Earth based on six criteria in the study area. This Google Earth-based method has the advantage of mapping landslides in inaccessible areas (Rabby and Li, 2019). This research also incorporated a 100-meter threshold-based accuracy assessment for Google Earth mapping (Galli et al. 2008). The accuracy of this mapping varies 69-88% based on the assessment of the two sites in the Chittagong Metropolitan Area (CMA) and Bandarban district. Five hundred forty-eight landslides were mapped using participatory field mapping (Samodra et al., 2018). In participatory field mapping, newspaper and government reports and published documents were used to determine where to carry out the field mapping. Then, the assistance of local people and stakeholders helped detect the actual location of landslides. Participatory field mapping helped identify and record landslides in urban areas, areas near road networks, and settlements. The combination of Google Earth mapping and participatory field mapping provided a detailed inventory with 730 landslides.

Landslide susceptibility mapping requires both presence (landslides) and absence (non-landslide locations) data (Zhu et al., 2019); however, the selection of absence-data is usually subjective. This research introduced MD-based absence-data sampling. MD values were calculated for 261 landslide locations using fifteen landslide causal factors, including slope, aspect, plan curvature profile curvature, geology, and distance from the road network. These MD values were compared with the Chi-square distribution to determine the critical value in determining the

space for absence-data sampling. The landslide susceptibility maps produced by the MD-based and slope-based absence-data sampling using the random forest model showed similar prediction performance at the test sites of the three Upazilas of Rangamati district, Bangladesh. However, the MD-based susceptibility map is more consistent and practically applicable. The slope-based susceptibility map classified more areas as high susceptibility zone, resulting in comparatively better accuracy but less consistency. In addition, the MD-based absence-data sampling is objective and statistically robust because it is based on a theoretical distribution and a specific confidence level.

Different from relatively stable factors, such as geology, slope aspect, plan curvature, and profile curvature. LULC is a dynamic factor affected by human activities (Reichenbach et al. 2014; Abedin et al. 2020). This research used the existing LULC of 2018, a simulated LULC (2028; also called the BAU scenario), and a proposed LULC to evaluate the impact of LULC on landslide susceptibility in the Rangamati municipality. The model produced satisfactory landslide susceptibility maps for all three LULC scenarios. The high susceptibility zone increases by 28.7% and 43.1% for planned and simulated LULC scenarios. It seems that although landslide susceptibility was not considered in the proposed LULC, the high susceptibility zone does not increase as high as for the BAU scenario. Neitherless, landslide susceptibility likely increases in both LULC scenarios.

5.2. Plans for the Future Work

This research established a criteria-based Google Earth mapping of landslides. Visual interpretation of Google Earth images was time-consuming and labour-intensive. In the future, automated methods can be developed in the Google Earth Engine to map landslides. The accuracy assessment used in this research can be applied to Google Earth Engine-based landslide mapping. The six criteria-based mappings can also be integrated into teaching to help the students develop knowledge on geomorphic analysis and visual interpretation of high-resolution images. High-resolution satellite images were not available in the study area. If funding is available, commercial, very high-resolution satellite images can be acquired in the future. Deep learning and machine learning-based methods can be used to detect landslides in satellite images.

Future studies can apply the MD-based absence-data sampling to various types of landslides in the world and evaluate the sensitivity of different confidence levels and casual factors. This research has demonstrated the impact of LULC change on landslide susceptibility and concentrated on the

change of all types of LULC. Future studies can assess the role of urban growth on landslide susceptibility. This study only considered LULC as a dynamic factor and treated other factors as static. In fact, climate change and its associated rainfall change are also dynamic factors. Future work is necessary to evaluate the impacts of climate change, especially the changing rainfall pattern, on the landslide susceptibility.

References

Abedin, J., Rabby, Y.W., Hasan, I. and Akter, H., 2020. An investigation of the characteristics, causes, and consequences of June 13, 2017, landslides in Rangamati District Bangladesh. *Geoenvironmental Disasters*, *7*(1), pp.1-19.

Ahmed, B., 2015. Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides*, *12*(6), pp.1077-1095.

Ahmed, B., Rahman, M.S., Sammonds, P., Islam, R. and Uddin, K., 2020. Application of geospatial technologies in developing a dynamic landslide early warning system in a humanitarian context: the Rohingya refugee crisis in Cox's Bazar, Bangladesh. *Geomatics, Natural Hazards and Risk*, *11*(1), pp.446-468.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P., 2008. Comparing landslide inventory maps. *Geomorphology*, *94*(3-4), pp.268-289.

Rabby, Y.W. and Li, Y., 2020. Landslide Susceptibility Mapping Using Integrated Methods: A Case Study in the Chittagong Hilly Areas, Bangladesh. *Geosciences*, *10*(12), p.483.

Rahman, M.S., Rahman, B.A.F.H.S. and T, M., 2016. Landslide inventory in an urban setting in the context of Chittagong Metropolitan Area, Bangladesh.

Rahman, M.S., Ahmed, B. and Di, L., 2017. Landslide initiation and runout susceptibility modeling in the context of hill cutting and rapid urbanization: a combined approach of weights of evidence and spatial multi-criteria. *Journal of Mountain Science*, *14*(10), pp.1919-1937.

Reichenbach, P., Mondini, A.C. and Rossi, M., 2014. The influence of land use change on landslide susceptibility zonation: the Briga catchment test site (Messina, Italy). *Environmental management*, *54*(6), pp.1372-1384.

Samodra, G., Chen, G., Sartohadi, J. and Kasama, K., 2018. Generating landslide inventory by participatory mapping: an example in Purwosari Area, Yogyakarta, Java. *Geomorphology*, *306*, pp.306-313.

Zhu, A.X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T. and Hong, H., 2019. A similarity-based approach to sampling absence-data for landslide susceptibility mapping using data-driven methods. *Catena*, *183*, p.104188.

Vita

Yasin Wahid Rabby comes from Bangladesh. He grew up in Dhaka, Bangladesh. He received a Bachelor of Sciences degree in Geography and Environment with a minor in Geology and Botany from the University of Dhaka Bangladesh in 2014. He published two papers during the undergraduate years and completed a thesis about "An Assessment of Microclimatic Variations: A Study in Dhaka City." He also completed an M.S. degree from the same university in Physical Geography in 2016. He completed a thesis about "Spatio Temporal Variability in Rainfall Over Bangladesh from 1980-2014". In 2016, he moved to the USA to enroll in the Department of Geography's doctoral program at the University of Tennessee, Knoxville. His fields of interest were geospatial data science, geomorphology, and hazard assessment. He was awarded the Doctor of Philosophy degree in May 2021. He worked as a graduate teaching and research assistant in the Department of Geography. He taught labs of various courses, including Introductory GIS, Intermediate GIS, Meteorology, World Regional Geography, and People and Environment. He has already published peer-reviewed research papers and actively participated in academic conferences. He has taken part in various voluntary and outreach activities. After graduation, Rabby seeks a faculty or research position in an educational institution.