Machine Learning-based Natural Terrain Landslide Susceptibility Analysis – A Pilot Study

H W M Li*, R H L Li, C C J Wong, F L C Lo

Geotechnical Engineering Office, Civil Engineering and Development Department, Government of HKSAR, Hong Kong SAR, China

*Corresponding author
doi: https://doi.org/10.21467/proceedings.133.8

ABSTRACT

Recently, the Geotechnical Engineering Office has initiated a pilot study on data-driven landslide susceptibility analysis (LSA) using a machine learning (ML) approach. A study area covering about one-fifth of the total natural hillside area of Hong Kong on and around the Lantau Island was considered. Three common tree-type ML classifiers: Decision Tree, Random Forest and XGBoost have been used. Conditioning factors (or features) including rainfall, geological and topography-related features were considered. In the study, the domain knowledge on natural terrain landslides in Hong Kong were critically incorporated into the susceptibility models through feature engineering to ensure that the resulted models are physically meaningful. In addition, an approach proposed to resolve the serious data imbalance problem, which is common in LSA, will be highlighted. Under this approach, the predicted probabilities of the positive class (i.e., landslide) can also be taken as a proxy to the landslide probability. This paper reports the methodology and key findings of this pilot study. The approach can be extended to cover other ML algorithms and features, and to a territory-wide scale with a view to enhancing the resolution and accuracy of the current susceptibility model of natural hillsides in Hong Kong.

Keywords: Stability Landslide Susceptibility, Feature Engineering, Machine Learning

1. Introduction

Much of the natural terrain in Hong Kong is steeply sloping with a surface mantle of weak saprolite, residual soil or colluvial deposits. These hillsides are susceptible to shallow rain-induced landslides with a typical depth of less than 3m. As part of its slope safety management system, the Geotechnical Engineering Office (GEO) has been conducting technical development work pertaining to landslide hazards. Landslide susceptibility analysis (LSA) has been one of the key areas of technical development work for improving our understanding of the nature and characteristics of natural terrain hillsides, their potential risk and approaches for risk management (Wong, 2009).

Landslide susceptibility refers to the spatial likelihood of a landslide occurring in an area on the basis of local terrain conditions (Brabb, 1984). While previous studies on landslide susceptibility for Hong Kong were mainly based on the data-driven analysis using conventional statistical approach (e.g. Evans & King, 1998; Ko & Lo, 2016), a few recent studies were based on the machine learning (ML) approach (e.g. Dai & Lee, 2002; Ng et al., 2021; Wang et al., 2021). The ML techniques are becoming popular to model complex landslide problems and starting to demonstrate promising the predictive performance compared to conventional methods (Tehrani et al., 2021). In light of this, the GEO carried out a pilot study to explore the potential of applying ML on natural terrain LSA. With reference to Ko & Lo (2016), landslides and conditioning factors covering the same period from year 1985 to 2008 were adopted. Under this pilot study, a study area covering about one-fifth of the total natural terrain areas in Hong Kong has been considered. Although it is a regional study, the methodology has been developed with a view to extending it to a territory-wide study. This paper presents the preliminary findings of this ML-based natural terrain LSA using three different ML algorithms.

Proceedings DOI: 10.21467/proceedings.133; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-957605-1-0

2. Previous Statistical-Based Natural Terrain Lsa for Hong Kong

The first territory-wide natural terrain LSA of Hong Kong conducted by the GEO is reported in Evans & King (1998). The territory was catergorised into five classes of susceptibility, with slope angle (13 classes) and geology (19 groups) accounted as the conditioning factors for landslide occurrence. The susceptibility classes were differentiated by one order of magnitude in terms of landslide frequency (i.e., 0.1 to 1 no. of landslide/km^2/year). This resolution was considered limited and insufficient in differentiating more vulnerable areas for risk management applications (Wong, 2003; Wong, 2009). Also, the effect of rainfall, a key contributory factor to landslide occurrence, was not considered in this study.

An updated territory-wide natural terrain susceptibility model was developed by Ko & Lo (2016) taking into account the effect of rainfall, together with the consideration of the enhanced landslide in the Enhanced Natural Terrain Landslide Inventory (ENTLI), and enhanced topography data from the territory-wide multi-return airborne Light Detection and Ranging (LiDAR) survey undertaken in 2010 (2010 LiDAR data) respectively. The three conditioning factors considered: slope gradient, solid geology and rainfall intensity were categorized into eight, three and six classes respectively and correlated to landslide susceptibility using conventional statistical approach. The analysis resulted in an improved resolution spanning across four to five orders of magnitude in terms of landslide density (no./km^2) (see Figure 1). In particular, natural terrain landslides were found to be highly sensitive to rainfall (two to three orders of magnitude between the lowest and highest rainfall classes, for one slope angle class). They further predicted the number of landslides that may occur in an anticipated rainfall event and generated a possible landslide frequency map (Ko and Lo, 2018). Based on the work by Ko & Lo (2016), a pilot study has been undertaken to explore the potential of applying ML on natural terrain LSA, taking advantage of the powerful performance of ML in data analytics. The following sections present the methodology and key findings of this ML-based study.

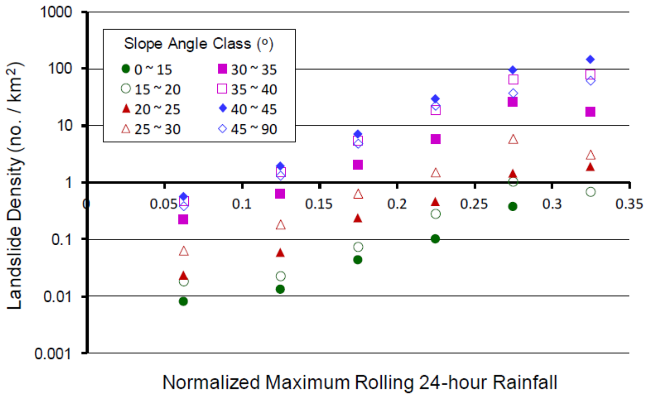

Figure 1: Year-based Rainfall Landslide Correlation by Ko & Lo (2016)

3. Machine Learning Algorithms

Since the first use of ML in the field of landslide studies in early to mid-2000s, a wide range of conventional ML and deep learning (e.g., neural networks) algorithms have been developed for classification and regression purposes. They have been used in various landslide studies yet there is still no consensus on which algorithm is the 'best' suited for predicting landslide prone areas (Dou et al., 2020). In this pilot study, ML algorithms which suit our purposes and the adopted approach of the study were identified based on the key factors below.

(i) interpretability of the algorithms,
(ii) balance between bias and variances,
(iii) suitability for handling correlated conditioning factors, and
(iv) computational efficiency.

Interpretability refers how easy is it to explain the results from the input data by a ML model, or to understand the patterns that models use to link to the training datasets (Ma et al., 2021). It is essential for detecting bias and debugging of the ML models. We also considered it is important for us to be able to explain the model predictions in combination with our professional knowledge (domain knowledge) on landslide occurrence. Balance between bias and variances refers whether the algorithm is able to form a predictive model that is generalized enough to give consistent yet accurate forward predictions. Algorithms which are prone to overfitting should be avoided. The ability of an algorithm to handle correlated conditioning factors, or more commonly referred as features in ML terminology, provide additional flexibility in selection of features and is thus more preferable. Lastly, computational efficiency is related to the time spent on the LSA and is taken as one of the considerations as well.

With reference to the above considerations, three tree-based ML algorithms, namely: Decision Tree, Random Forst and XGBoost were chosen. Decision Tree algorithm (Breiman et al., 1984) is known as one of the most commonly used algorithms in the studies of similar nature. Despite it is a less robust algorithm and sensitive to the predictive data, it is adopted for its computational efficiency and high interpretability to facilitate the understanding of the other two algorithms. Random forest (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016) are tree-based ensemble learning algorithms. With different ensemble methods adopted, their performance has been much enhanced in terms of robustness and generalizability. While tree-based ensemble algorithms are widely recognized to achieve excellent results compared to other ML algorithms, Ma et al. (2021) in particular remarked that Random Forest algorithm offers robust performance for accurate susceptibility mapping with only a small number of adjustments required before training the model. On the other hand, the performance of XGBoost has been widely recognized in a number of ML and data mining challenges (e.g., Kaggle competitions). XGBoost is one of the Gradient Boosting algorithms. While Gradient Boosting become popular very recently such that they are less routinely used in LSA, it has been reported to be able to improve the accuracies of ML models for landslide susceptibility analyses (Merghadi et al., 2020).

The architecture of the adopted algorithms are not elaborated in this paper. Readers may refer to the original papers of the algorithms for more details.

4. Modelling Approach and Feature Engineering

4.1. Modelling Approach

The LSA in this pilot study was treated as a grid-based binary classification problem in ML. In other words, given the set of condition factors possessed by a grid, the ML classifier predicts the landslide occurrence within the cell as a binary dependent variable comprising positive value (with landslide) or negative value (without landslide) only. The same approach was adopted by similar studies in Hong Kong (Dai & Lee, 2001; Ng et al., 2021; Wang et al., 2021). A grid size of 5m x 5m was adopted as in the previous work by Ko & Lo (2016). Reichenbach et al. (2018) also remarked that grid-based approach is the most common type of mapping units for LSA modelling.

One of the major challenges commonly encountered in applying binary classification in LSA is the sample bias due to the highly imbalance dataset, as there is always a scarce proportion of positive value grids within a study area. The ratio of positive value to negative value grids the dataset for this pilot study area is in the order of 1:30,000. Such imbalances can cause a model to be biased towards classifying the susceptible area as safe (i.e., negative value), jeopardizing the accuracy of the minority class prediction (the class of interest in our study). With a view to improving the binary classification result of the minority class, data-level techniques which refer to selecting a 1:1 ratio (or other ratio as appropriate) of landslide data points to non-landslide data points using different sampling techniques were commonly adopted (Dai & Lee, 2001; Ng et al., 2021; Ma et al., 2021). However, there is concern that sampling of data would bias the predicted probabilities of a classifier, resulting in a significantly high proportion of false positive when applying the classifier trained and tested using sampled datasets to an unsampled domain. This manifest as a substantial over-prediction of landslide potential when practically applying such models to forward prediction of the landslide potential of the entire study areas.

In view of the above, this pilot study adopted a different ML approach to handle an imbalance dataset (referred as the adopted approach in this paper). Under this approach, while the analysis was still handled as a binary classification problem, no data sampling was applied to avoid biasing the predicted class probabilities of a classifier. The issue of sample bias was overcome by taking the 'predicted class probabilities' (e.g. the probability of having a landslide), instead of the 'predicted classes' (i.e. with or without landslide), as the key prediction result obtained from the ML classifier. Since the predicted probabilities of the classifier were not biased by data sampling, the predicted probability of the positive class [P(+ve)] could be taken as a proxy to the landslide probability directly. In addition, the information loss in model training can be minimized as the entire dataset except those saved for validation and testing can be utilised without sampling. Similar approach was adopted in Xiao & Zhang (2021) in forecasting the number of man-made slope failures in response to rainstorms with machine learning technique for slope-based analysis.

4.2. Feature Engineering

The nature and number of features adopted in ML modelling vary significantly among different literatures on LSA. While the introduction of redundant or irrelevant features may create noise that decreases the overall predictive capability of the models, no universal agreement on the principle of selecting relevant features among the literatures could be found. In particular, some of the features adopted in the literatures lack physical relevance with landslide occurrence. As a result, it is preferable to identify appropriate features through understanding their roles on landslide occurrence and the adoption of conventional engineering which involves a substantial amount of prior knowledge (Reichenbach et al., 2018; Ma et al., 2021).

In view of the above, a framework which ensures the quality, and the statistical and physical relevance of the features was adopted in selecting proper features for inclusion in this pilot study.

4.2.1. The Feature Selection Framework

Feature selection is the process of reducing the dimensionality of input variables and creating summary measures to encapsulate the information in the entire dataset. Domain knowledge would be used in the process to extract the characteristics and attributes from raw data. Under the feature selection framework, potential features were assessed against the criteria given in Table 1. In particular, while ML belongs to algorithmic modelling which provides prediction based on the available data only and treat the data mechanism as an unknown (Tehrani et al., 2021), we emphasised on the inclusion of our domain knowledge for the development of physically meaningful ML models.

Table 1 Feature Selection Criteria

Criterion	Consideration
(i) quality of the feature datasets	• good spatial and temporal coverage, resolution and accuracy of the feature data • crucial to ensuring the performances of ML models are not adversely affected by the quality of their training data • avoid underrepresentation due to scarcity of the available data

Proceedings DOI: 10.21467/proceedings.133
ISBN: 978-81-957605-1-0
Series: AIJR Proceedings
ISSN: 2582-3922

(ii) statistical correlation between the feature and landslide occurrence	• prevent the introduction of less relevant or redundant features to the ML model, as they create noise that decreases the overall predictive capability of the models • facilitate understanding of the data structure • aid the data preprocessing by revealing under representative data (e.g. the response of the hillsides under heavy rainfall with a low probability of occurrence, characteristics of the terrain covering a small land area) • assessed by means of descriptive analytics (see Table 2)
(iii) consistency of the statistical correlation with domain knowledge on landslide susceptibility	• allows the incorporation of domain knowledge, past experience and expert judgment on landslide susceptibility to the ML model by reviewing the consistency of the correlation observed in Criterion (ii) against the existing engineering knowledge on landslide occurrence

A feature selection priority matrix (the matrix) shown in Figure 2 was created for assessment against Criteria (ii) and (iii) in this pilot study. Potential features that fall within Quadrant 1 would have higher priority to be included in the susceptibility models, follow by those falling in Quadrant 2. Features found in Quadrants 3 and 4 reveal statistical correlations that do not tally with the existing engineering knowledge and should be well considered and tested with due consideration of the data representativeness before inclusion.

Further elaborations on the application of the matrix is given based on three topography-related example features: slope gradient, profile curvature and aspect. The vertical axis of the matrix is determined with reference to Table 2, which summarises the correlations of the example features with the density of past landslide occurrences ($\delta_{Landslide}$, no./km^2) and their distribution of area. Territory-wide data of Hong Kong instead of the study area solely were used with a view to extend the pilot study to territory-wide in future.

Strong statistical correlation with $\delta_{Landslide}$ increases by about five times from 30^o to 45^o is observed from slope gradient. The increase of $\delta_{Landslide}$ with slope gradient is attributable to its effects on the balance of stabilizing and destabilizing forces, and thus the overall stability of a slope. Areas steeper than 45^o possess lower $\delta_{Landslide}$ as they are more rocky or composed of denser soil, having a higher stabilizing force. This feature is thus placed near the high ends of both the vertical and horizontal axes in the matrix.

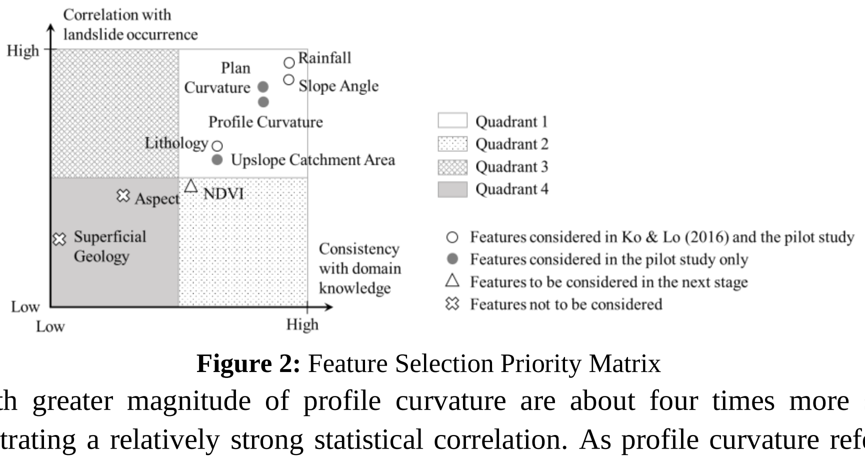

Figure 2: Feature Selection Priority Matrix

Areas with greater magnitude of profile curvature are about four times more susceptible to landslide, illustrating a relatively strong statistical correlation. As profile curvature refers the rate of change of slope gradient along the vertical directions, it can be considered as a proxy to the break in slope which is landslide related (Ho & Roberts, 2016). This feature is thus placed in Quadrant 1 of the matrix, at a less extreme position as compared to slope gradient.

Aspect shows certain degree of statistical correlation with landslide, with the south or southeast aspects being two times more susceptible than the north. Nonetheless, this correlation cannot be justified based on domain knowledge. As such, aspect falls within Quadrant 4 and is not considered in this pilot study.

Table 2 Area and Landslide Density Distribution of Example Features

Slope Gradient (deg)	Area (%)	$\delta_{Landslide}$ (no./km^2)	Profile Curvature	Area (%)	$\delta_{Landslide}$ (no./km^2)	Aspect	Area (%)	$\delta_{Landslide}$ (no./km^2)
0 - 15	12.5%	1.20	$\leq$ -7	2.1%	77.75	N	12.2%	20.19
15 - 20	12.0%	2.19	-7 - -5	2.7%	61.65	NE	12.3%	25.75
20 - 25	17.4%	4.61	-5 - -3	7.5%	43.08	E	12.7%	31.35
25 - 30	22.6%	12.57	-3 - -1	20.7%	25.06	SE	12.6%	38.63
30 - 35	19.3%	42.21	-1 - 1	36.1%	17.29	S	12.5%	39.05
35 - 40	10.2%	102.36	1 - 3	18.7%	27.32	SW	12.5%	34.88
40 - 45	4.1%	141.68	3 - 5	6.3%	52.28	W	12.6%	30.82
45 - 90	2.0%	95.36	5 - 7	2.6%	76.90	NW	12.5%	25.37
			$\geq$ 7	3.4%	58.08			

4.2.2. The Selected Features

Table 3 summarizes the features identified based on the feature selection framework. As can be seen, compared with the three basic features considered by Ko & Lo (2016) (i.e. rainfall, lithology, slope gradient), this study considered three additional features (i.e. plan curvature, profile curvature and upslope catchment area). These features are briefly described as follows.

Table 3: Summary of Features Considered in this Study

Feature	Rainfall	Lithology	Slope gradient	Plan Curvature	Profile Curvature	Upslope Catchment Area
Ko & Lo (2016)	✓	✓	✓	-	-	-
This Study	✓	✓	✓	✓	✓	✓

4.2.2.1. Rainfall

The natural terrain landslides in Hong Kong were characterized as rainfall-induced. Since 1980s, the GEO and the Hong Kong Observatory (HKO) have installed a total of more than 120 automatic rain gauges across Hong Kong, with an average density of about one rain gauge per 10 km^2. In particular, the GEO rain gauges (90 nos.) capture real-time data which would be transmitted to the servers automatically at up to 1-minute intervals. This network of rain gauges provides a reasonably good spatial and temporal coverage of rainfall data across Hong Kong.

Year-based rainfall intensities quantified in terms of normalised maximum rolling rainfall (NMRR) was adopted in this pilot study. NMRR is determined by normalizing the maximum rolling rainfall as recorded at a location with the mean annual rainfall of the same location of a 30-year period from year 1977 to 2006. The normalisation of rainfall intensity is a common approach to better characterise extremity or anomalies of the rainfall (Ko, 2005; Ko & Lo, 2016). The statistical correlation of 24-hour and 4-& 24-hour NMRR with landslide susceptibility was thoroughly studied in Ko & Lo (2016), which concluded that landslide occurrence is highly sensitive to rainfall with a strong statistical correlation up to five orders of magnitude observed. With reference to Ko & Lo (2016), rolling durations of 24-hour was considered in this pilot study.

4.2.2.2. Topography-related Features

Topography-related features refer to the list of features which could be determined with reference to the topography of the study area. A total of four topography-related features (i.e. slope gradient, plan curvature, profile curvature and upslope catchment area) were identified for incorporation in this pilot study. The data of these features obtained from the 2010 LiDAR data. The 0.5m x 0.5m digital terrain model (DTM) developed was first resampled to form a 5m x 5m DTM, and then converted into the feature dataset using ArcGIS applications. Given LiDAR data shows promising results in producing high resolution DTM that can 'see through' vegetation, topography-related feature dataset derived from territory-wide LiDAR survey results are of good quality fulfilling feature selection Criterion (i).

Statistical correlations of two to six folds are observed from these features. In terms of physical significance, slope gradient plays a significant role on the overall stability of a slope. Plan and profile curvatures are related to the mass-wasting and runoff processes. The upslope catchment area indicates the amount of flow that would be concentrated to the grid in event of precipitation.

4.2.2.3. Lithology

Lithology has been adopted in both generations of the territory-wide natural terrain landslide susceptibility analysis by Evans & King (1998) and Ko & Lo (2016). The lithology is related to the engineering properties of the soils derived from the parent rocks and is thus considered to be physically relevant to the landslide potential. The lithology of the study area was categorised into three main groups (namely intrusive, volcanic, and sedimentary) with reference to the 1:20,000 solid and superficial geology maps of Hong Kong (https://www.cedd.gov.hk/eng/publications/geo/hong-kong-geological-survey/index.html). The same categorization was adopted in Ko & Lo (2016).

4.2.3. The Landslide Data

In this pilot study, landslide data as recorded in the ENTLI was adopted. The ENTLI provides year-based landslide information which were collected through the interpretation of available high-flight (≥ 2,400m altitude) and low-flight aerial photos (< 2,400m altitude). As compared with the reported landslides or field-mapped landslides which are commonly adopted in other LSA, landslide data based on ENTLI provided a more complete picture of landslide occurrence over the study area which is not biased by the accessibility of the landslide locations. On the other hand, the temporal resolution of the data was limited by the frequencies of aerial photo-taking and interpretation. Year-based LSA was considered as a result. In the dataset, grids containing the crowns of the landslides were identified as the landslide area and denoted as '1'; remaining grids were considered as non-landsliding area and denoted as '0'.

5. Machine Learning-Based Natural Terrain Landslide Susceptibility Analysis

5.1. The Pilot Study Area

The pilot study area comprises the natural terrain areas of the Lantau Island, as well as those of the adjacent outlying islands as indicated in Figure 3. It is 130 km^2 by area, over 30% of the study area is steeper than 30^o, with the elevation varied from sea level to 930m above sea level. It is mainly underlain by volcanic and intrusive rock, with a small area of sedimentary rock. There were over 6,100 recent natural terrain landslides recorded within the study area in the ENTLI. The study area has experienced intense rainfall in 1993 and 2008, with the 24-hour maximum rolling rainfall of over 500 mm and 600 mm respectively. The rainstorm on 7 June 2008 alone has resulted in over 2,500 natural terrain landslides. Given the high variabilities in topography-related and rainfall data available within the pilot study area, as well as its rich history of past landslides, it is considered as an ideal study area for this pilot study.

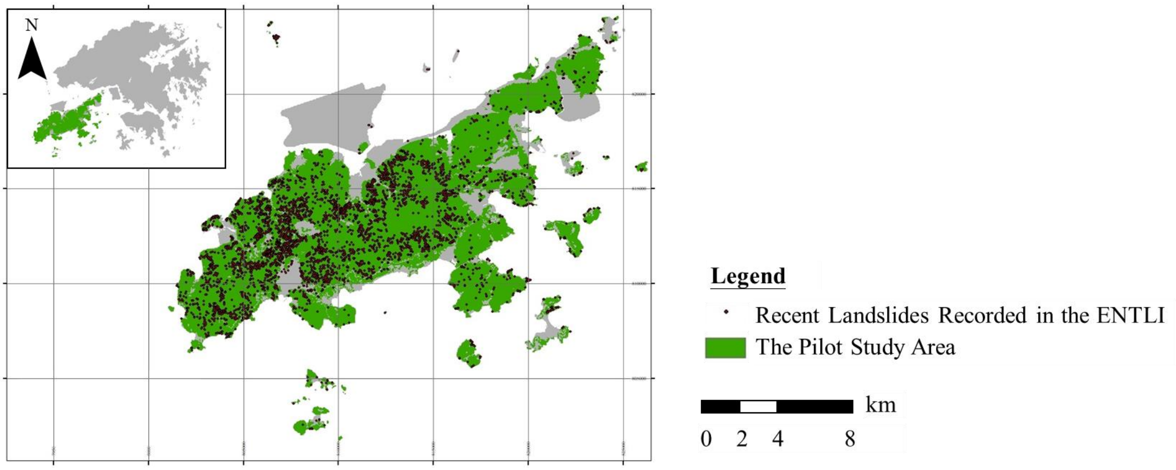

Figure 3: Extent of the Study Area and Recent Landslides in Enhanced Natural Terrain Landslide Inventory (ENTLI)

5.2. The Workflow

The workflow of this ML-based analysis mainly comprises data preprocessing and resampling, model construction and performance evaluation stages.

5.2.1. Data Preprocessing and Resampling

With a grid-based approach adopted in this pilot study, the entire pilot study area discretized into about 5.2 million numbers of 5m x 5m grids, each of which contains 24 years (year 1985 to 2008) of rainfall and landslide data on top of the geological and topography-related features. Under the adopted approach, most of the data in grids were used for either the construction or the evaluation of the ML models. Given the amount of data to be handled, the model construction and evaluation works of this pilot study were carried out on web service platform using python programming language.

Data preprocessing refers to the preparation of data for model construction. Key actions include the cleansing of data, the encoding of categorical data, as well as the resampling of data for model training and evaluation. Data cleansing forms part of the feature engineering works, which involves the removal of null or undesirable data from the dataset to ensure only representative and unbiased data are fed into the models. For instance, data associated with rainfall intensities beyond the range of 0.025 to 0.3 for 24-hour NMRR were removed since only a limited portion of the pilot study area had encountered these extreme rainfall intensities in rare events, such that the associated data would not be representative enough for incorporation. Data points associated with the extreme values of plan and profile curvatures were discarded for similar reasons. The encoding of categorical data involved the lithological data only, with one-hot encoding adopted.

The resampling of the dataset along with the workflow of this pilot study is illustrated in Figure 4, the dataset was resampled into training dataset, validation dataset and testing dataset. The testing dataset comprised 1) all the data from years 1993 and 2007, and 2) 10% of the data randomly selected from the remaining 22 years in a stratified manner. Stratified random sampling is a commonly adopted sampling technique in which the data is divided into smaller groups or strata and then randomly selected from each of the strata by the same proportion. The data were stratified based on landslide occurrences in this pilot study such that the ratio of landsliding to non-landsliding data in each of the dataset could be maintained.

The two sets of testing data are referred as Testing Data 1 (TD1) and Testing Data 2 (TD2) respectively. While TD2 pertained the type of testing data commonly adopted in other similar studies,

Proceedings DOI: 10.21467/proceedings.133
ISBN: 978-81-957605-1-0
Series: AIJR Proceedings
ISSN: 2582-3922

TD1 comprised data that possess unseen rainfall patterns during the model construction such that it served as a more stringent test which tested the models' ability in making forward predictions. Of note, the intensity of rainfall in year 1993 is one of the highest one among the 24 years, whilst that in year 2007 is a moderate one. The remaining data served as training and validation datasets.

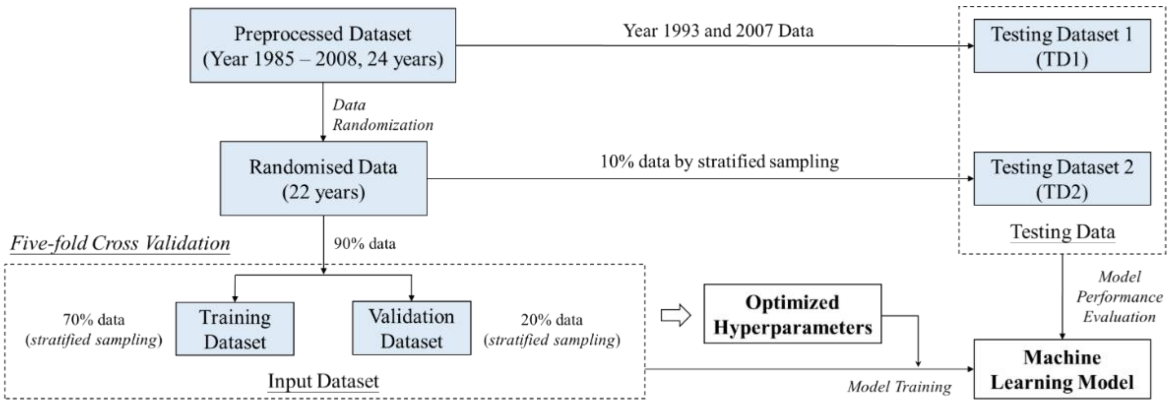

Figure 4: Resampling of Data in the Pilot Study

5.2.2. Model Construction

Construction of the ML model mainly involves the optimization, or tuning, of the hyperparameters. Hyperparameters are parameters that control the learning process of a ML model, the optimal set of hyperparameters to be used varies by cases as it is dependent on the algorithm and dataset involved. Although some ML-based studies adopt the default values of hyperparameters, Tehrani et al. (2021) remarked that hyperparameter tuning plays a significant role on the performance and the predictive ability of a ML model. The tuning of hyperparameters is a process of trial and error, which can be done in either systematically (e.g grid search) or randomly. For each set of the hyperparameters considered in this pilot study, their performances were assessed using five-fold cross-validation. A five-fold cross validation involved the key steps below:

(1) shuffling of dataset in a random manner
(2) splitting of the shuffled dataset into nine groups in a stratified manner
(3) from (2), select seven groups of the split data as training dataset to fit the model
(4) evaluate the trained model using the remaining two groups of split data, i.e., the validation dataset, with reference to the area under the receiver operating characteristic (ROC) curve
(5) repeat (3) and (4) for a total of five times

Each of the analysis cases were trained and evaluated with the set of preprocessed data based on the same workflow. Scikit learn packages were used for the implementation of the Decision Tree and Random Forest algorithms, whereas the XGBoost package was adopted for the XGBoost algorithm.

5.2.3. Performance Evaluation

Performances of the ML models in this pilot study were evaluated using the Area Under Curve (AUC) of the receiver operating characteristic (ROC) curves (see Figure 5). An ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) for different classification thresholds, such that the pre-determination of the threshold is not required. A higher ROC AUC value indicates a better performance of the model over the whole range of the classification threshold. This evaluation metric is chosen since it does not require a predetermined classification threshold, which is non-trivial given the modelling approach adopted in this pilot study, to define the splitting of the model predictions in binary classification.

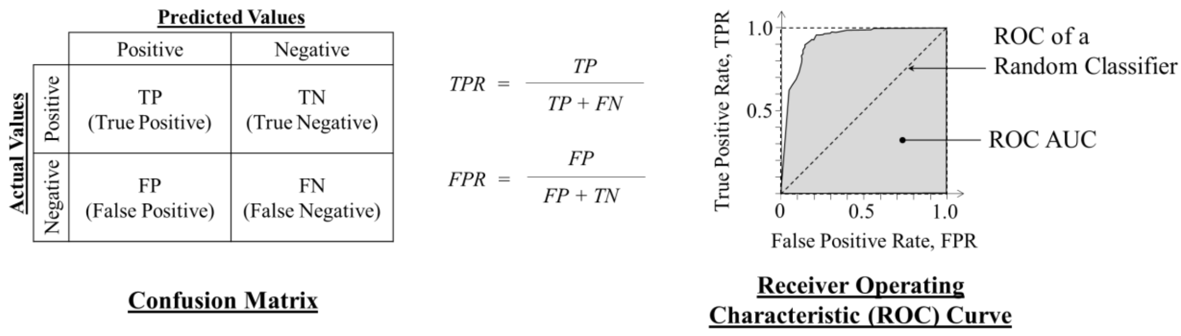

Figure 5 Definition of ROC AUC and Confusion Matrix

The ROC AUC of the ML models assessed based on the different data are summarised in Table 4. A set of ML models considering the features adopted in Ko & Lo (2016) only are also included as reference case.

Table 4 ROC AUC of the Predictive Models

Evaluation Data	Training Data			Testing Data 1 (TD1)			Testing Data 2 (TD2)		
ML Algorithms	Decision Tree	Random Forest	XGBoost	Decision Tree	Random Forest	XGBoost	Decision Tree	Random Forest	XGBoost
Pilot Study	0.9674	0.9965	0.9808	0.8764	0.9083	0.9149	0.9443	0.9670	0.9732
Reference case	0.9571	0.9881	0.9660	0.8770	0.8836	0.8867	0.9499	0.9593	0.9627

On top of the ROC AUC evaluation, Tehrani et al. (2021) suggested to further examine the accuracy of a ML model by validating the areal extent of each susceptibility class against the density distribution of landslides in the landslide inventory. A model is accurate when the landslide density ratio increases moving from low to high susceptibility classes, and when the high susceptibility classes cover small extent of areas only. Table 5 validates the XGBoost model which gives the highest ROC AUC for both TD1 and TD2 among the models tested as an illustration. The pilot study area is divided into eight groups based on the P(+ve) as predicted by the model in a log cycle. The (a) distribution of actual landslides from year 1985 to 2008 as recorded in the ENTLI, and the (b) areal distribution of the study area in each year is tabulated by group. The actual landslide density, $\delta_{Landslide}$, in each of the groups is calculated as dividing (a) by (b).

Table 5 Distribution of Actual Landslide Occurrences and Area by Predicted P(+ve)

Group	I	II	III	IV	V	VI	VII	VIII
Predicted P(+ve)	10^{-8}-10^{-7}	10^{-7}-10^{-6}	10^{-6}-10^{-5}	10^{-5}-10^{-4}	10^{-4}-10^{-3}	10^{-3}-10^{-2}	10^{-2}-10^{-1}	10^{-1}-1
Area (km^2)	117.6	1486.8	1056.8	318.8	100.9	19.1	1.4	0.004
	(3.8%)	(48%)	(34%)	(10%)	(3.3%)	(0.6%)	(0.05%)	(0.0001%)
Landslide no.	1	30	167	484	1268	1586	817	24
	(0.02%)	(0.69%)	(3.82%)	(11.1%)	(29.0%)	(36.2%)	(18.7%)	(0.55%)
$\delta_{Landslide}$ (no./km^2)	0.009	0.020	0.158	1.518	12.57	83.09	564.3	5,962.7
Actual P(+ve) *	2.13x10^{-7}	5.04x10^{-7}	3.95x10^{-6}	3.80x10^{-5}	3.14x10^{-4}	2.08x10^{-3}	1.41x10^{-2}	1.49x10^{-1}

*Actual P(+ve) refers the actual probability of landslide, which is the product of actual landslide density and grid size.

Proceedings DOI: 10.21467/proceedings.133
ISBN: 978-81-957605-1-0
Series: AIJR Proceedings
ISSN: 2582-3922

6. Discussion

6.1. Accuracy

Key observations from the performance evaluation results are summarised below.

(a) The ROC AUC based on training data of all cases are over 96%, indicating that the ML models fit the training data very well.

(b) The ROC AUC based on testing data TD1 and TD2 are over 87%. The maximum ROC AUC were up to 91.5% and 97.3% respectively for the XGBoost models. The ROC AUC based on TD1 is obviously lower than that based on TD2 for all of the cases as the former served as more stringent test on the models' abilities in making forward predictions. All in all, the ROC AUC values achieved reveal that all of the ML models are able to make fairly accurate predictions.

(c) XGBoost and Random Forest models perform better than Decision Tree models based on the testing data. The results are also shown in Table 4. As compared with the reference case, the introduction of additional features improved the performance of the ML models, the effect is more obvious when the models were tested with TD2.

(d) In Table 4, about 85% (3,695 out of 4,377 nos.) of the landslides fall within 4% of area of the highest landslide susceptibility (Susceptibility Classes V to VIII), demonstrating that the ML model is giving fairly accurate susceptibility predictions.

6.2. Predicted Probability vs Actual Probability

Under the adopted approach, the predicted P(+ve) is directly taken as the predicted landslide probability. This section validates this assumption. Figure 6 plots the actual landslide probability [Actual P(+ve) in Table 5] of the pilot study area from year 1985 to 2008 by P(+ve) as predicted by the XGBoost model. A linear relationship with a gradient of unity is observed, indicating that the two quantities are close to each other. As such, the predicted P(+ve) of the ML model under the adopted approach provides a fairly realistic indication of, and can been taken as a proxy to, the predicted landslide probability for practical applications.

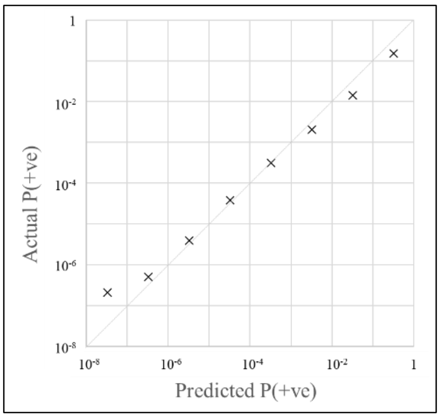

Figure 6: Actual Landslide Probability by Positive Classification Scores (XGBoost model)

6.3. Spatial Resolution

The spatial forecast of landslide susceptibility models is often presented as susceptibility maps. Each of the grids on the map represents the reclassified or calibrated landslide occurrence prediction of the location covered. Table 6 compares the range of the landslide probability of the entire pilot study area as predicted by the XGBoost model under rainfall intensities corresponding to the mean normalized 24-hour NMRR intensities of 24-hour NMRR Classes I to V in Ko & Lo (2016). The range of the landslide probability based on an XGBoost model considering the three features adopted in Ko & Lo (2016) only is also included for reference (the reference model). With the additional topography-related features considered in the former model, it differentiates the landslide susceptibility of terrain with a higher spatial resolution by two to three orders of magnitude for each of the rainfall intensity classes considered.

Figure 7 shows an extract of the landslide susceptibility maps near the Tai O area as predicted by the reference model and the XGBoost model under a hypothetical constant rainfall scenario with 24-hour NMRR Class IV to illustrate the difference. Again, the XGBoost model is able to distinguish landslide susceptibility with a much higher resolution as compared with FS1-24hr-XGB.

Table 6 Range of Predicted Probability of the Pilot Study Area

24hr NMRR		Rainfall Class I	Rainfall Class II	Rainfall Class III	Rainfall Class IV	Rainfall Class V
XGBoost Model	Min	9.89×10^{-9}	1.17×10^{-8}	4.17×10^{-8}	1.35×10^{-7}	2.17×10^{-7}
	Max	1.57×10^{-3}	1.63×10^{-3}	1.34×10^{-2}	9.59×10^{-2}	2.71×10^{-1}
Reference Model	Min	2.12×10^{-7}	2.54×10^{-7}	6.35×10^{-7}	3.45×10^{-6}	9.61×10^{-6}
	Max	2.04×10^{-5}	3.36×10^{-5}	2.10×10^{-4}	2.68×10^{-3}	2.56×10^{-3}

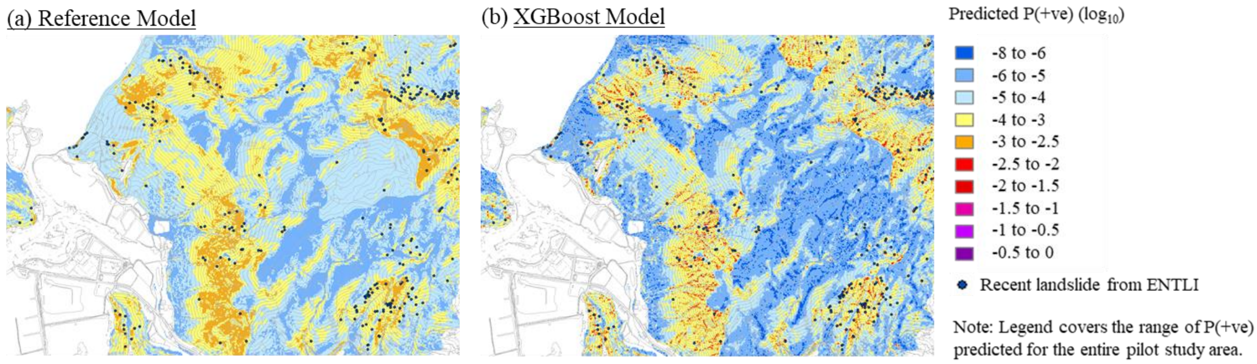

Figure 7 Landslide Susceptibility Map of the Tai O Area (24h-NMRR Class IV)

7. Concluding Remarks

Over the years GEO has been conducting technical development work on LSA for natural hillsides. This pilot study is carried out to explore the potential improvement to the existing landslide susceptibility model of natural terrain that can be brought about by the application of ML analysis. It is different from the other ML-based LSA on two aspects: 1) the adoption of a different modelling approach instead of data sampling to tackle the issue of acutely imbalanced dataset, and 2) the placing of emphasis on the incorporation of domain knowledge throughout entire workflow of the study.

The adopted modelling approach is proven to work fairly well, with the ML models giving accurate susceptibility predictions which can be taken as a proxy to landslide probability. ML-models allow a systematic way to include additional features for LSA. The results of the study show that the resolution of the susceptibility map is enhanced by two to three orders of magnitude upon the introduction of three critically assessed additional features. The degree of improvement to the spatial resolution of the landslide susceptibility map is similar for the range of rainfall intensity considered.

However, a fine balance should be struck between the predictive performance and the interpretability of the model. While being powerful, the ML algorithms learn the association between landslide occurrences and the set of features in various manner without considerations on the physical mechanism of slope failure behind. As such, the use of ML does not guarantee better susceptibility models that is physically meaningful unless it is applied with the input of sound professional knowledge. In this pilot study, we have demonstrated the introduction of domain knowledge to machine learn-based models through critical feature engineering works, proper selection of suitable algorithms, and detailed assessment of the model performances.

As a pilot study, the conducted analyses focused on the group of the most promising features and algorithms, based on rainfall and landslide data up to year 2008 only. Before the study is expanded to a territory-wide scale, we believe the models can be further enhanced on various aspects. Suitable additional features fulfilling the same feature selection framework will be introduced in a step-wise manner with a view to maximizing the amount of information gain while maintaining the feature space of the dataset in a reasonable dimension. Additional ML algorithms may also be considered. As this pilot study considered rainfall and landslide data up to year 2008 only, the dataset will be expanded to cover data of nearer years for testing of the ML models. Retraining of the ML models will also be carried out if necessary, especially when data associated with high rainfall intensity become available. Rainfall intensity characterised in different rolling durations may be considered. Scale effect and effect of post-landslide topography may also be explored.

8 Declarations

8.1 Acknowledgements

This paper is published with the permission of the Head of the Geotechnical Engineering Office and the Director of Civil Engineering and Development, the Government of the Hong Kong Special Administrative Region, China.

8.2 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Brabb, E.E. 1984. Innovative approaches to landslide hazard mapping. Proceedings of 4th International Symposium on Landslides, 1, 307-324.

Breiman, L., Friedman, J.H., Olshen, R.A.& Stone, C.J. (1984). Classification and Regression Tree. Wadsworth, New York.

Breiman, L. (2001). Random Forest. Machine Learning, 45, pp. 4-32 (2001).

Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA: 785-794.

Dai, F.C. & Lee, C.F. (2001). Terrain-based mapping of landslide susceptibility using a geographical information system: a case study. Canadian Geotechnical Journal, 38, 911-923.

Dou, J., Yunus, A.P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B.T.,

Yamagishi, Ho., (2020). Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Science of The Total Environment, 720, 137320.

Evans, N.C. and King, J.P. (1998). The Natural Terrain Landslide Study - Debris Avalanche Susceptibility (TN 1/98), Geotechnical Engineering Office, Hong Kong.

Ko, F.W.Y. & Lo, F.L.C (2016). Rainfall-based landslide susceptibility analysis for natural terrain in Hong Kong - A direct stock-taking approach. Engineering Geology. 215: 95-107.

Ko, F.W.Y. & Lo, F.L.C (2018). From Landslide Susceptibility to Landslide Frequency: A territory-wide study in HK. Engineering Geology. 242: 12-22.

Li, H.W.M., Lo, F.L.C., Wong, T.K.C & Cheung, R.W.M. (2022). Machine Learning-Powered Rainfall-Based Landslide Predictions in Hong Kong - An Exploratory Study. Not yet published.

Ma, Z., Mei, G. & Piccialli, F. (2021). Machine Learning for Landslides Prevention: A Survey. Neural Computing and Applications. 33(17), 10881–10907.

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., Pham, B.T., Bui, D.T., Avtar, R., Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225.

Ng, C.W.W., Yang, B., Liu, Z.Q., Kwan, J.S.H., Chen, L. (2021). Spatiotemporal modelling of rainfall-induced landslides using machine learning. Landslides, 18: 2499-2514.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihri, M., Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-Science Reviews 2018, 180: 60-91.

Tehrani, F.S., Calvello, M., Liu, Z.Q., Zhang, L., Lacasse, S. (2021). Machine learning and landslide studies: recent advances and applications, Not yet published.

Wang, H., Zhang, L., Lau, H., He, J., Cheung, R.W.M. (2021). AI-powered landslide susceptibility assessment in Hong Kong. Engineering Geology, 288, 106103.

Wong, H.N. (2003). Natural Terrain Management Criteria – Hong Kong Practice and Experience. Thirty Years of Slope Safety Practice in Hong Kong. Geotechnical Engineering Office, Hong Kong.

Wong, H.N. (2009). Rising to the challenges of natural terrain landslides. HKIE Geotechnical Division Annual Seminar 2009, Hong Kong, 15–53.

Xiao, T. & Zhang, L. (2021). Storm-based forecasting of man-made slope failures using machine learning. Proceedings of the 20th International Conference on Soil Mechanics and Geotechnical Engineering, Sydney 2021.