

Research Paper

Landslide identification using machine learning

Haojie Wang^a, Limin Zhang^{a,*}, Kesheng Yin^a, Hongyu Luo^a, Jinhui Li^b^a Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China^b Department of Civil and Environmental Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

ARTICLE INFO

Keywords:

Landslide risk
Landslide identification
Machine learning
Deep learning
Big data
Convolutional neural networks

ABSTRACT

Landslide identification is critical for risk assessment and mitigation. This paper proposes a novel machine-learning and deep-learning method to identify natural-terrain landslides using integrated geodatabases. First, landslide-related data are compiled, including topographic data, geological data and rainfall-related data. Then, three integrated geodatabases are established; namely, Recent Landslide Database (ReLD), Relict Landslide Database (ReLD) and Joint Landslide Database (JLD). After that, five machine learning and deep learning algorithms, including logistic regression (LR), support vector machine (SVM), random forest (RF), boosting methods and convolutional neural network (CNN), are utilized and evaluated on each database. A case study in Lantau, Hong Kong, is conducted to demonstrate the application of the proposed method. From the results of the case study, CNN achieves an identification accuracy of 92.5% on ReLD, and outperforms other algorithms due to its strengths in feature extraction and multi dimensional data processing. Boosting methods come second in terms of accuracy, followed by RF, LR and SVM. By using machine learning and deep learning techniques, the proposed landslide identification method shows outstanding robustness and great potential in tackling the landslide identification problem.

1. Introduction

Landslide identification plays an important role in landslide risk assessment and management (Guzzetti et al., 2012; Zhang et al., 2012; Chen et al., 2016; Naidu et al., 2018). With the advent of the remote sensing technology, landslides can be identified through visual interpretation of both remote sensing images and topographic surfaces (Haneberg et al., 2009; Xu, 2015). Although the visual interpretation has high identification accuracy, the process is time-consuming and labor-intensive. Hence, automated or semi-automated methods for landslide identification based on remote sensing techniques are highly sought after in recent years.

Current studies on landslide identification are mainly based on optical images using pixel-based or object-oriented methods, and the digital terrain model (DTM) is often used as auxiliary data for such analysis (Guzzetti et al., 2012). Barlow et al. (2003) combined optical images and digital elevation model (DEM) derivatives to identify translational landslide scars using object-oriented methods. Mckean and Roering (2004) constructed high-resolution DEMs and used their derivatives to identify bedrock landslides near Christchurch, New Zealand. Martha

et al. (2010) used a combination of spectral, shape and contextual information to identify landslides, and further used multi-temporal images to identify historical landslides (Martha et al., 2012), both are based on object-oriented methods. Keyport et al. (2018) further investigated the feasibility of pixel-based and object-oriented landslide identification methods using very high-resolution images.

Machine learning and deep learning techniques have been proven to be a powerful and promising tool in many geotechnical applications (Zhang et al., 2015; Lary et al., 2016; Zhang and Goh, 2016; Papaioannou and Straub, 2017; Ching and Phoon, 2018; Li et al., 2018, 2019; Lo and Leung, 2019) as well as in landslide identification. Moosavi et al. (2014) compared pixel-based methods implemented by artificial neural network (ANN) and support vector machine (SVM) with object-oriented approaches in producing landslide inventories. Van Den Eeckhaut et al. (2012) used data segmentation and SVM to identify forested landslides with DTM derivatives. Li et al. (2015) used random forests (RF) and SVM to identify forested landslides in the Three Gorges area of China with DTM derivatives based on object-oriented methods. Ding et al. (2016) attempted automatic recognition of landslides using CNN and texture change detection with pre- and post-landslide optical images in

* Corresponding author.

E-mail address: cezhangl@ust.hk (L. Zhang).

Peer-review under responsibility of China University of Geosciences (Beijing).

<https://doi.org/10.1016/j.gsf.2020.02.012>

Received 27 September 2019; Received in revised form 10 December 2019; Accepted 22 February 2020

Available online 10 March 2020

1674-9871/© 2020 China University of Geosciences (Beijing) and Peking University. Production and hosting by Elsevier B.V. All rights reserved. This is an open

access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

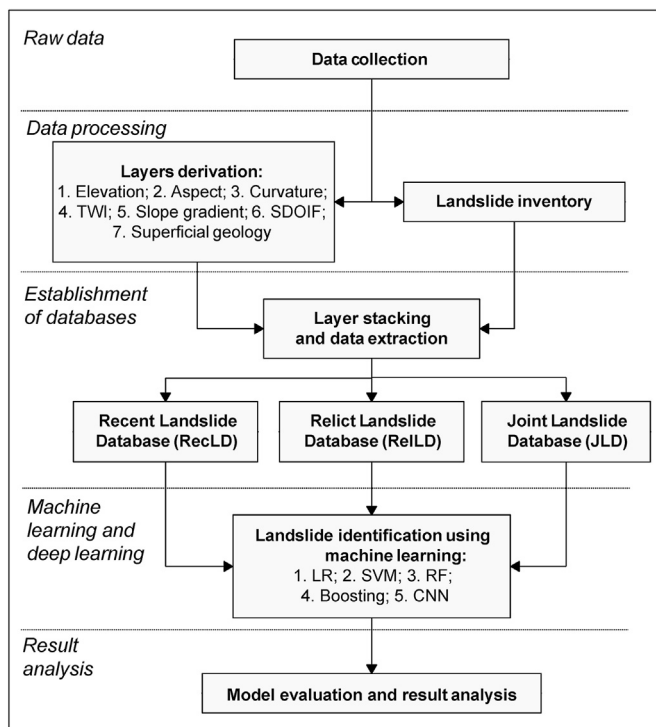


Fig. 1. Proposed landslide identification method using machine learning.

Shenzhen, China. Ghorbanzadeh et al. (2019) evaluated the performance of ANN, SVM, RF and convolutional neural network (CNN) in detecting landslide areas in Rasuwa District, Nepal, with optical images and DEM derivatives.

Although efforts have been made to develop efficient landslide identification methods, several problems remain unsolved concerning the application of machine learning and deep learning towards landslide identification:

- (1) Current studies mainly focus on identifying landslides with available optical images. The identification of relict landslides is barely explored.
- (2) The literature pays more attention to the identification process itself; the impacts of different landslide types on identification performance remain unclear.
- (3) The literature relies heavily on the optical remote sensing images to conduct landslide identification. The latest high-resolution DTM can precisely capture minor terrain differences; yet the potential of DTM-dominant landslide identification using machine learning and deep learning has not been exploited.

The main objective of this paper is to propose an integrated landslide identification method which is able to identify both relict and recent landslides from DTM using machine learning and deep learning. A block-based data extraction method is used to establish different types of landslide databases. On each database, various machine learning and deep learning models based on LR, SVM, RF, boosting and CNN are trained and compared to evaluate their performance. A case study of Lantau Island is worked out to validate the proposed method with the assistance of a Recent Landslide Database (RecLD), a Relict Landslide Database (RelLD) and a Joint Landslide Database (JLD).

2. Landslide identification using machine learning and deep learning

In this study, machine learning and deep learning techniques are adopted (1) to evaluate the performance of the proposed machine-

learning and deep-learning based method for landslide identification, and (2) to investigate the performance of the proposed method when applying to identify different types of landslides. The overarching methodology is summarized in Fig. 1 and will be introduced step by step in the paper.

2.1. Raw data

Raw data that is needed can be divided into two categories: environmental data and landslide inventory. The environmental data characterize the overall environment or settings of landslides. Landslides, especially relict landslides, are hard to identify from optical remote sensing images due to the presence of a dense vegetation cover or other reasons. A better way would be to identify landslides from a DTM which truly reveals the surface change without being affected by the vegetation cover. Owing to the advance of modern remote sensing, more and more high-resolution DTMs become available. Even intricate terrain features can be captured using such high-resolution DTMs. In addition, data that relate to the slope failure mechanisms should also be gathered. For example, to evaluate the stability of a slope, the factor of safety (FOS) of a slice of an infinite slope, when the slip surface is below the groundwater table, is given by (Griffiths et al., 2011; Wang et al., 2019):

$$\text{FOS} = \frac{(H\gamma\cos^2\beta - u)\tan\phi' + c'}{H\gamma\sin\beta\cos\beta} \quad (1)$$

where H is the depth of the soil layer to the potential failure surface; β is the slope inclination; γ is the total unit weight of the soil above the failure surface; u is the pore water pressure at the failure surface; c' and ϕ' are the effective cohesion and effective friction angle of the soil at the failure surface, respectively. Based on Eq. (1), environmental data that represent or relate to these six parameters (H , β , γ , u , c' and ϕ') will be needed. Overall three general types of environmental data are needed:

- (1) topographic data: Data such as elevation, curvature and slope gradient are essential to describe the terrain and such data can be derived from a DTM. In addition, elevation may have a relation to H as the weathering conditions are likely to be similar in a small interval of elevation;
- (2) geological data: The regional geological conditions are critical for landslide identification, which also have impacts on the γ , c' and ϕ' of the superficial soil deposits;
- (3) rainfall-related data: Rainfall is one of the most common triggers of landslides (Gao et al., 2018) and it can greatly affect the pore water pressure u .

Hence, in the data collection step, three important types of data should be considered: DTMs, geology maps, and rainfall-related data.

Landslide inventories are another essential part of raw data. Landslides can be classified into two kinds: relict and recent. Recent landslides are those that clearly occurred within the time scale of available aerial photographs; relict landslides are those that occurred earlier than the time scale of the available aerial photographs (GEO, 1996; Maunsell-Fugro Joint Venture and GEO, 2007). Both kinds of landslide data can be obtained through aerial photograph interpretation (API) and topographic surface interpretation.

2.2. Data processing

The raw data will be processed in two stages. The first stage is to derive landslide-related predictors from the raw data. With reference to Eq. (1) and landslide susceptibility studies (Reichenbach et al., 2018), four geomorphological predictors and one hydrological predictor are derived from the DTM, namely, elevation, aspect, curvature, slope gradient and topographic wetness index (TWI). Besides, the superficial geology and rainfall-related data are also included as two predictors. A

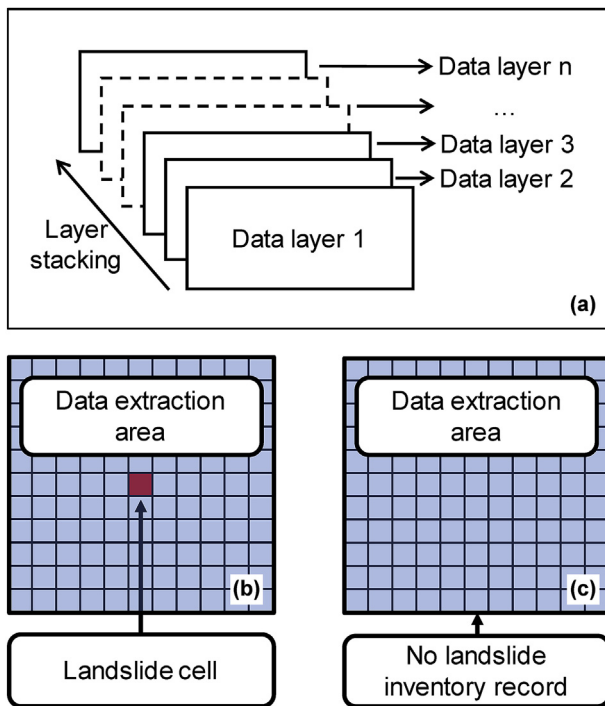


Fig. 2. Diagrams for (a) layer stacking, (b) data extraction for positive and (c) negative samples.

step duration orographic rainfall intensification factor (SDOIF) is chosen to quantify the relationship between the rainfall intensity and terrain surface features (AECOM and Lin, 2015; Gao et al., 2017). It is worth mentioning that the predictors chosen in the proposed method are a reference choice rather than a compulsory one. Predictors utilized in different cases should be determined according to their own specific conditions. Two principles should be followed when choosing the predictors: (1) the predictors should be closely related to either the triggering and failure mechanisms of slopes or the ground features of the landslide locations; (2) the predictors should be quantifiable and available. The second stage is to transform the data to a machine-readable format. Coupled with the GIS platform, the data of the targeted research area is transformed into raster with reasonable cell size. For each predictor, a data layer is formed to store the data of the designated area.

2.3. Establishment of databases

The next step is to establish databases for learning, which includes three components: layer stacking, data extraction and establishment of databases. As shown in Fig. 2a, by stacking the data layers that represent different predictors, the original one-dimensional data of each layer are stacked to form an n -dimensional dataset that contains the conjunct data of all layers. Moreover, as this study aims to investigate the performance difference of the same machine learning or deep learning algorithm on different types of landslides, three different landslide inventories are formed; namely, the Relict Landslide Inventory (ReLI), the Recent Landslide Inventory (ReLI) and the Joint Landslide Inventory (JLI). Note that the JLI contains both the relict and recent landslides. Combining the n stacked layers of landslide predictor data, each landslide inventory can then be mapped to the corresponding locations for data extraction. The locations of all landslides are then obtained.

As all data layers of a designated area have gone through rasterization in the data processing step, each landslide record is then assigned to a cell according to its location. The cell which contains the landslide record is defined as a landslide cell. Shown in Fig. 2b, the data located in the

surrounding areas of the landslide cells will be extracted to generate positive samples for learning. The size of the extraction area can be specified according to actual demand. Fig. 2b shows an extraction area of 11 by 11 cells. Note that a 'positive sample' here means that there is a recorded landslide in the data extracted area of the sample and the extracted data include the whole n -dimensional data within the surrounding area. Meanwhile, shown in Fig. 2c, according to the corresponding landslide inventory, random areas where no recorded landslide exists are picked to generate negative samples. For a particular landslide inventory, the number of positive samples is given. The number of negative samples is set to be identical to that of the positive ones to avoid skewed data problem (James et al., 2013).

With the extracted positive and negative samples from different landslide inventories, the samples are combined to form three databases; namely, the Relict Landslide Database (ReLD), the Recent Landslide Database (ReLD) and the Joint Landslide Database (JLD). For each established database, the data which is originally extracted from the same data layer needs to be normalized as they may have different orders of magnitude. For example, the elevation may reach nearly thousand-meter level while the slope gradient ranges from 0° to 90° . The inconsistent magnitude of data layers may cause slow convergence or convergent failure of adopted machine learning algorithms (James et al., 2013). Hence, normalization is conducted to each dimension of the established databases.

2.4. Machine learning and deep learning algorithms

Having established the databases, the next step is to train various machine learning and deep learning models and conduct landslide identification. In this study, landslide identification is a binary classification problem, each sample is given a prediction of either positive or negative using the trained models. Four types of machine learning algorithms and one deep learning algorithm are chosen to evaluate the feasibility of machine learning and deep learning in landslide identification.

2.4.1. Logistic regression (LR)

As one of the most widely-used classification algorithms, logistic regression shows satisfactory performance with relatively low computational cost (James et al., 2013). A logistic function for multiple variable logistic regression can be written as:

$$P(\mathbf{X}) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_m}} \quad (2)$$

where $\mathbf{X} = (X_1, \dots, X_m)$ are m predictors, the set of β are the corresponding learning parameters, and $P(\mathbf{X})$ is the probability of being positive given the predictors \mathbf{X} . The maximum likelihood method is often used to estimate the learning parameters.

2.4.2. Support vector machine (SVM)

The support vector machine is another popular classification approach that was developed in the 1990s. It is considered as one of the most adaptable algorithms for its good performance in various settings. A support vector machine uses kernels to enlarge the feature space and quantify the similarity of two observations, therefore to capture the non-linear decision boundary for classification (James et al., 2013). For example, Eq. (3) describes a popular kernel named radial kernel:

$$K \left(x_i, x_j \right) = \exp \left(-\lambda \sum_{j=1}^m (x_{ij} - x_{ij})^2 \right) \quad (3)$$

where x_{ij} and x_{ij} are the i th pair of observations of the j th predictor, m is the number of predictors, λ is a tuning parameter which accounts for the smoothness of the decision boundary and K stands for the kernel function.

2.4.3. Random forest (RF)

Random forest is essentially a tree-based method, which has a reliable and good prediction performance by combining a large number of decision trees to yield a single consensus prediction. The major feature of random forest is that it is not allowed to consider a majority of the available features at each split of the tree (Friedman et al., 2001). For example, a common practice for choosing the number of splits is:

$$m' = \sqrt{m} \tag{4}$$

where m' is the number of predictors at each split and m is the number of all predictors.

2.4.4. Boosting

As a branch of the ensemble methods, boosting is a way of combining the performance of a number of weak classifiers to produce a powerful “committee”, so it is regarded as a strong classifier. As an example, the prediction given by Discrete AdaBoost is described as:

$$F(x) = \text{sign} \left(\sum_1^M c_m f_m(x) \right) \tag{5}$$

where $f_m(x)$ is a weak classifier that produces either positive or negative prediction, c_m is a coefficient calculated by learning weights, M is the number of weak classifiers, the *sign* function here returns either a positive or negative prediction and $F(x)$ is the corresponding prediction. By combining several models, the boosting method can achieve a better prediction performance compared with a single model. The main procedure of boosting is to fit a sequence of weak learners (e.g., discriminant analysis, k nearest neighbors, decision tree, etc.) to weighted versions of training data (Friedman et al., 2000). In this study, three popular boosting algorithms are selected to implement the landslide identification; namely Discrete AdaBoost, LogitBoost and Gentle AdaBoost.

2.4.5. Convolutional neural network (CNN)

As one of the most popular deep learning algorithms, convolutional neural networks have gained much attention for its remarkable contribution in computer vision (Goodfellow et al., 2016). A typical CNN is composed of four key components: convolutional layers, activation layers, pooling layers and fully connected layers. Built with these layers, many well-designed CNN structures have been proposed in many fields of study. In this study, each sample with different data layers is like a special ‘picture’ with multiple channels. The classification of high dimensional data is also the strength of the CNN.

2.5. Model evaluation

In this study, seven performance evaluation indices are used to assess the performance of each model. In order to calculate these seven indices, concepts of four kinds of predicted samples for classification learning need to be clarified:

- (1) true positive (TP): the predicted class is positive and the prediction agrees with the actual class;
- (2) false positive (FP): the predicted class is positive while the prediction disagrees with the actual class;
- (3) true negative (TN): the predicted class is negative while the prediction agrees with the actual class; and
- (4) false negative (FN): the predicted class is negative and the prediction disagrees with the actual class.

Seven performance indices, i.e., accuracy, precision, recall, specificity, false positive rate (FPR), F_1 score and Matthews correlation coefficient (MCC) are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

$$\text{FPR} = \frac{FP}{TN + FP} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

$$F_1 \text{ score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{12}$$

The accuracy quantifies the percentage of samples which are correctly predicted on a certain database. The specificity measures the proportion of actual negatives that are correctly identified. A model with a high specificity is more talented in classifying negative samples. FPR is the probability of false alarm. The precision evaluates the fraction of true positive samples among all predicted positive samples, while the recall quantifies the fraction of true positive samples among all actual positive samples. A high precision means that the model has a high probability to give a correct positive sample classification. While the recall explains how sensitive the model is towards identifying the actual positive samples; in other words, the recall quantifies the probability of detecting actual positive samples. As the precision and recall evaluate different aspects of the model, an index which combines both is also used. The F_1 score is the harmonic mean of the precision and recall, where an F_1 score reaches its best value at 1 and worst at 0. A more comprehensive index, the MCC that returns a value between -1 and 1 is also used. A coefficient of 1 represents a perfect prediction; 0 , no better than random prediction; -1 , total disagreement between prediction and observation. Both F_1 score and MCC are measures of model accuracy. As accuracy, F_1 score and MCC are all proportional to the model performance, this study proposes a simple combined index, overall accuracy (OA), to quantify the overall performance of a model:

$$\text{OA} = \text{Accuracy} + F_1 + \text{MCC} \tag{13}$$

The model with the highest OA will be chosen as the best model for landslide identification in any area of interest.

Besides, predictor importance (PI) is an important index to evaluate which predictor contributes the most for the model training. The magnitude of PI for different models trained by different algorithms may be inconsistent and the predictor importance ranking for different models may also be slightly different. Hence, an index called averaged relative predictor importance (ARPI) is proposed to assess the general predictor importance on multiple algorithms. For example, the ARPI for predictor i with q algorithms can be given by:

$$\text{ARPI}(i) = \frac{1}{q} \sum_{j=1}^q \frac{\text{PI}(i,j)}{\sum_{i=1}^m \text{PI}(i,j)} \tag{14}$$

where $\text{PI}(i, j)$ is the importance of the i th predictor trained by the j th algorithm, m is the total number of predictors, and q is the number of algorithms.

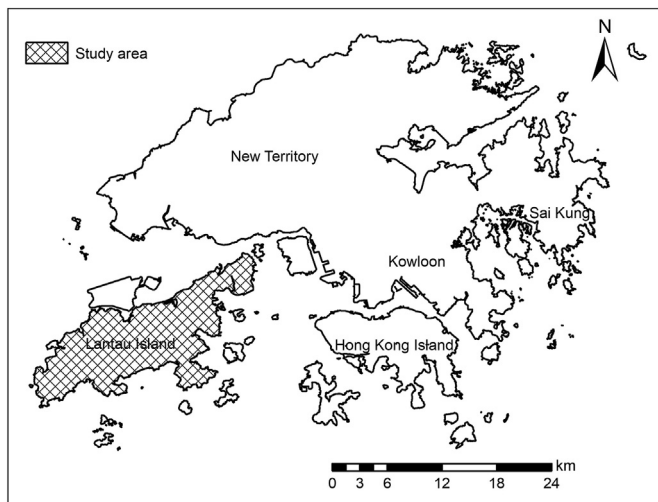


Fig. 3. Location of the study area.

3. Landslide identification on Lantau Island, Hong Kong

3.1. Study area

As shown in Fig. 3, the study area covers the entire Lantau Island. With a total area of around 147 km², Lantau Island is the largest outlying island located at southwestern Hong Kong. Due to the steep terrain, there is only a small amount of flat ground near the seaside. Inactive human activities cause a relatively complete natural state around Lantau Island. In Fig. 4, the main bed rocks on Lantau Island are weathered volcanic rocks and granitic rocks, which are usually covered with younger alluvial and colluvial materials. The oldest rocks are sandstone and siltstone, and these sedimentary rocks usually have smaller outcrops. Mixed forest grows at the foot of the hillslope, while thick shrubs and weeds develop in the middle of slopes. Outcrops of bedrock usually appear on the peak or in areas with slope angles greater than 40°. With a subtropical monsoon climate, the area is warm and dry in winter but hot and humid in summer. Lantau Island is affected by frequent high intensity storms and typhoons, making it a landslide-prone area (Fig. 5) and giving this

area high scientific value of landslide studies.

3.2. Data

A landslide inventory “Enhanced Natural Terrain Landslide Inventory (ENTLI)” has been established by Geotechnical Engineering Office (GEO, 1996; Maunsell-Fugro Joint Venture and GEO, 2007). The ENTLI contains both recent and relict landslide records from 1924 to 2009 in the study area and the spatial distribution of landslides is visualized in Fig. 6. Table 1 summarizes the landslide inventory for the study area. There are 5810 recent landslides and 20,884 relict landslides. The recent landslides are identified through visual interpretation of aerial photographs and are classified into three kinds: channelized landslides, open hillslope landslides and costal landslides. The relict landslides that occurred earlier than the time scale of available aerial photographs are identified based on API and terrain characteristics and are classified into classes A, B or C, with different interpretation confidence. Similar to the recent landslides, there is an independent class for costal landslides. As mentioned in the methodology, three landslide inventories are formed based on the ENTLI data; namely, the Relict Landslide Inventory (ReLLI), the Recent Landslide Inventory (ReCLI) and the Joint Landslide Inventory (JLI).

Apart from the landslide inventory, DTM, superficial geology and rainfall-related data are also prepared for this case study. Table 2 summarizes seven predictors for this case study. These predictors are classified into three types:

- (1) topographic predictors: elevation (Fig. 7a), aspect (Fig. 7b), curvature (Fig. 7c), topographic wetness index (TWI) (Fig. 7d) and slope gradient (Fig. 7e);
- (2) geological predictor: superficial geology (Fig. 7f); and
- (3) rainfall-related predictor: 24-h step duration orographic intensification factor (SDOIF) (Fig. 7g).

First, a high-resolution DTM of the study area with raster cell size of 2 m × 2 m is utilized to derive data layers using GIS. Specifically, with the digital elevation of the terrain surface, three widely used slope characteristics (i.e., slope gradient, curvature and slope aspect) are calculated for each cell. Also, a hydrological parameter TWI is calculated to consider the hydrological flow path on the terrain surface:

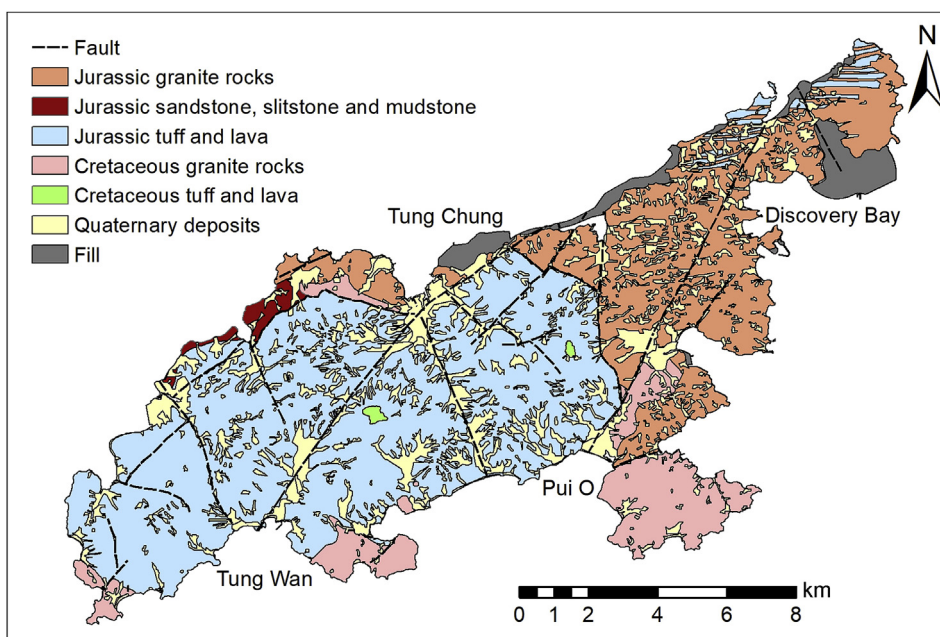


Fig. 4. Simplified geological map of the study area.



Fig. 5. Landslides near Ngong Road, Southwestern Lantau, 9th October 2011 (Source: Google Earth; coordinates: 22°14'50"N, 113°53'55"E).

$$TWI = \ln(a/\tan b) \tag{15}$$

where a is the local upslope area draining through a certain point per unit contour length and $\tan b$ is the local slope.

Second, the superficial geological map of the study area is extracted from the 1: 20,000 geo-map from Geology Survey of Hong Kong. In total, as shown in Fig. 7f, seven types of surface geomaterials (i.e., Jurassic granite rocks; Jurassic sandstone, siltstone and mudstone; Jurassic tuff and lava; Cretaceous granite rocks; Cretaceous tuff and lava; Quaternary deposits; and fill) are distributed over the terrain surface of the study area. Note that for non-Quaternary deposit and non-fill area, the rock mineral is usually subject to heavy weathering in the shallow surfaces. As the lithology is a categorical variable, a dummy variable approach is applied. To be specific, the following treatment is conducted:

For each cell j within the study area,
 For each geomaterial type i ($i \in [1, 2, 3, 4, 5, 6, 7]$), create a dummy variable :

$$x_i = \begin{cases} 1, & \text{if the } i\text{-type geomaterial agree with the cell material} \\ 0, & \text{if the } i\text{-type geomaterial disagree with the cell material} \end{cases}$$

 Endfor
 Endfor

(16)

where the seven geomaterial types correspond to seven types of superficial geomaterials.

In addition, a factor called SDOIF is also included in this study to consider the effect of orographic amplification on rainfall. Based on regional rainfall data, the SDOIF over time period t for a given cell j can be calculated by:

$$f_t(j) = \frac{r_t(j)}{r_{0,t}(j)} \tag{17}$$

where $f_t(j)$ is the average SDOIF over time t for cell j , $r_t(j)$ is the total rainfall containing orographic influence for cell j , and $r_{0,t}(j)$ is the rainfall without the orographic influence. The SDOIF data is extracted from AECOM and Lin (2015) with a resolution of 5 km × 5 km; Kriging is then used to derive the SDOIF data layer for the entire study area and the result is shown in Fig. 7g.

3.3. Preparation of databases

After the preparation of the seven data layers of predictors, a layer stacking process is conducted to integrate the data into a multiple-layer dataset (Fig. 7h). In this step, the data from the whole study area are organized in a tensor form with dimensions of 12,637 × 9737 × 13. As the dummy variable approach is applied to the superficial geology layer, the original superficial geology data layer is extended to 7 dimensions. Including the other six data layers, there are 13 layers in total.

After the establishment of the fundamental dataset, the three landslide inventories (i.e., RelLI, RecLI and JLI) are mapped on the study area individually and data extraction is carried out in each case. As mentioned

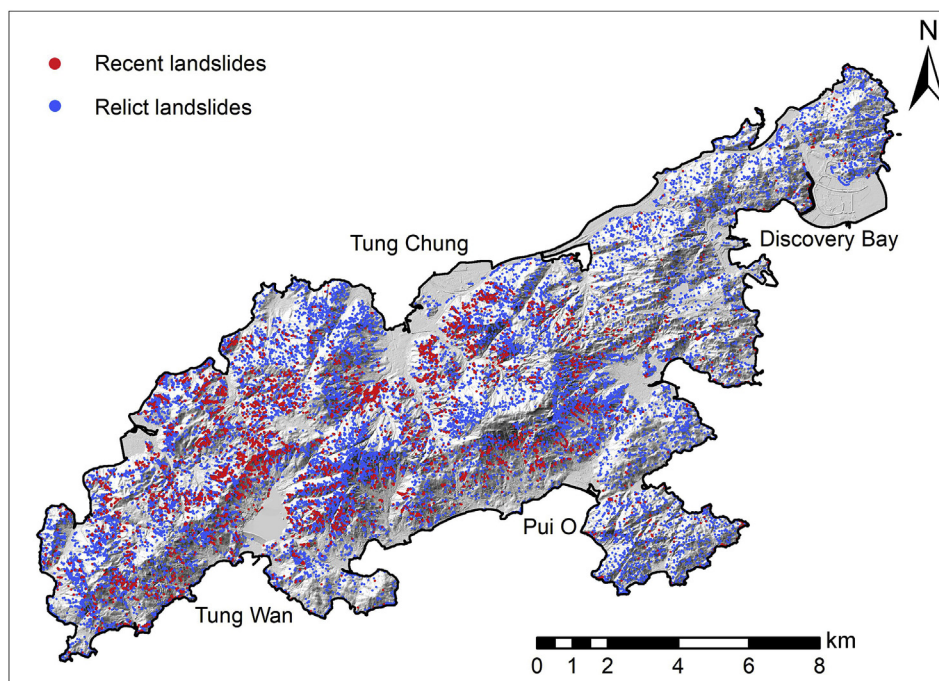


Fig. 6. Spatial distribution of recent and relict landslides in ENTLI on a shaded relief map of the study area.

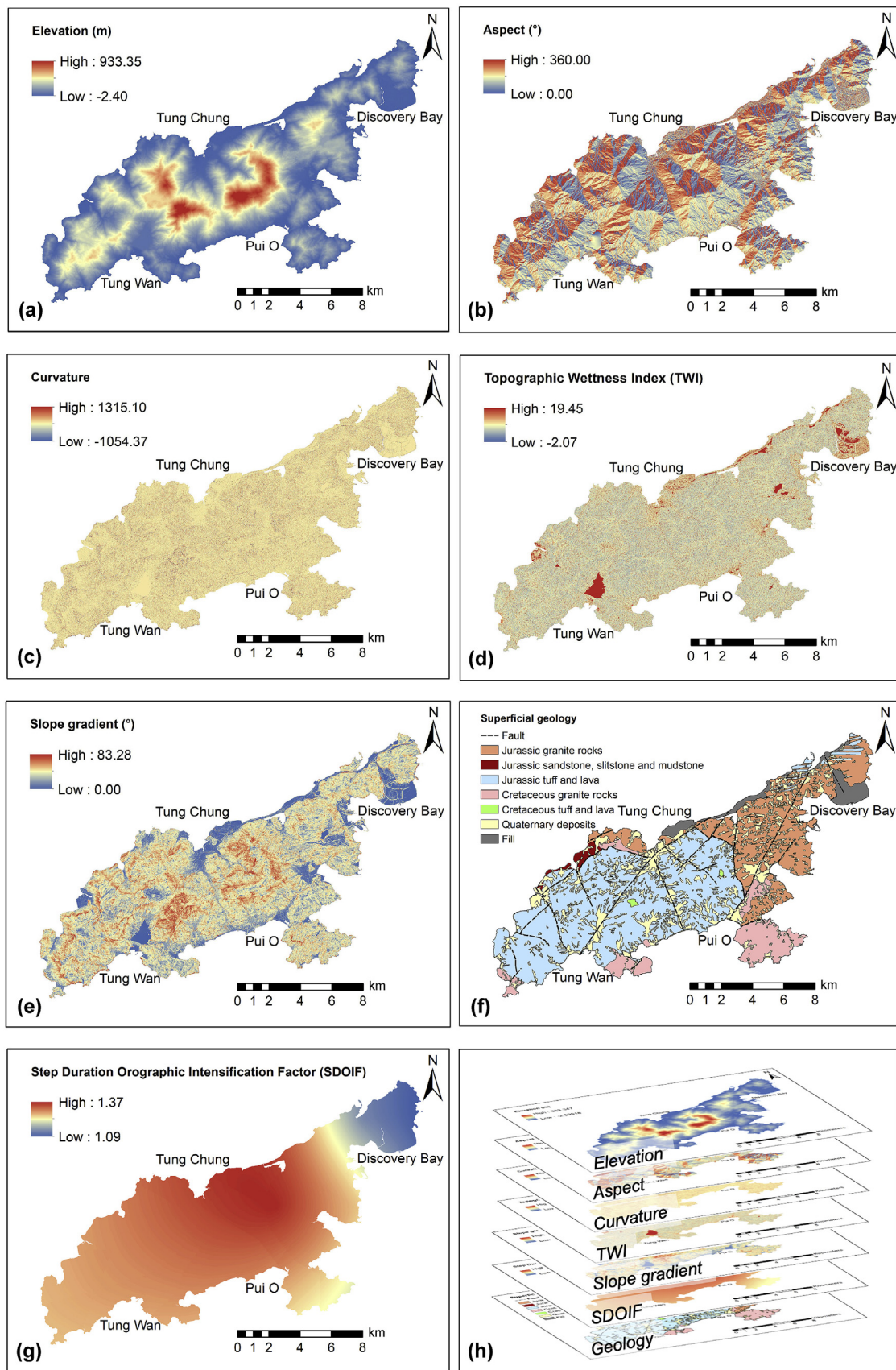


Fig. 7. Multiple data layers: (a) elevation; (b) aspect; (c) curvature; (d) TWI; (e) slope gradient; (f) superficial geology; (g) SDOIF; and (h) all layers stacked.

Table 1
Summary of 26,694 landslides in Enhanced Natural Terrain Landslide Inventory (ENTLI) within the study area.

Recent landslides		Relict landslides	
Class	No. of landslides	Class	No. of landslides
Channelized	2593	A	6480
Open hillslope	3144	B	9007
Costal	73	C	4529
		Costal	868
Subtotal	5810	Subtotal	20,884

before, the whole research area is rasterized into 2 m × 2 m cells. After the landslide mapping, landslide cells can then be confirmed according to the mapped landslide locations. Then, the size of data extraction area, viz., sample size, needs to be determined. In order to ease the computational load of model training in later stages while capturing as many terrain features as possible, the sample size is set to be 22 m × 22 m as the widths of 93.03% of the landslide records in this case study are smaller than 22 m. Hence, an area of 22 m × 22 m is utilized to conduct data extraction for each positive or negative sample by referring to Fig. 2b and c. Positive samples are extracted from the areas that contain the landslide cells which means there are inventory-recorded landslides, and negative samples of the same number are extracted randomly from areas where no landslide records are found in the inventory. Examples of a positive sample and a negative sample are displayed in Fig. 8. Noted that, for the convenience of demonstration, the superficial geology layer is not processed with the dummy variable approach. In each positive sample, there usually exists a clear landslide scar (as marked with red lines in Fig. 8); while in a negative one, no landslide characteristics can be observed. After the data extraction, three landslide databases are then established by integrating the samples: Recent Landslide Database (ReLD), Relict Landslide Database (ReLD) and Joint Landslide Database (JLD). Very few costal landslides have a distance less than 22 m from the coastline; these records are excluded from the databases.

3.4. Landslide identification based on landslide databases ReLD, ReLD and JLD

ReLD, ReLD and JLD contain 11,552, 41,320 and 52,872 samples in total, respectively. The positive and negative sample ratio is 50/50 for each landslide database.

As described in the methodology section, five types of machine learning algorithms, viz. LR, SVM, RF and Boosting (Discrete AdaBoost, LogitBoost and Gentle AdaBoost), and one deep learning algorithm, viz. CNN, are applied to each database separately. As each algorithm can be trained and tuned with a variety of parameters and settings, configurations which are typical for their general applications are chosen. For the deep learning algorithm CNN, the depth of CNN often needs to be tuned for specific applications. Hence, two CNNs structures with different

Table 2
Summary of predictors.

Topographic, geology and rainfall-related predictors	Description	Source, scale/ resolution	Data type	Data summary			
				Minimum	Average	Maximum	Standard deviation
Elevation (m)	Digital elevation of the terrain surface	DTM, 2 m	Continuous	-2.40	187.84	933.35	166.19
Slope gradient (°)	Angle of the slope inclination	DTM, 2 m	Continuous	0.00	24.49	83.28	12.20
Curvature	Curvature of the slope, indicating concave or convex surfaces	DTM, 2 m	Continuous	-1054.37	1.78	1315.10	18.04
Aspect (°)	Exposition of the slope	DTM, 2 m	Continuous	0.00	181.13	360.00	103.40
Topographic wetness index (TWI)	An index to quantify the topographic control on hydrological process	DTM, 2 m	Continuous	-2.07	2.93	19.44	2.12
Superficial geology	Lithology of the surface material	Geo-map, 1: 20,000	Categorical				
Step duration orographic intensification factor (SDOIF)	An index to quantify the amplification of orography on rainfall	AECOM and Lin (2015), 5 km × 5 km	Continuous	1.08	1.29	1.37	0.06

numbers of layers are built in this case study. Detailed layer settings of the two CNNs are displayed in Fig. 9. One with 6 layers is denoted as CNN-6 (i.e., two convolutional layers, two max-pooling layers and two fully connected layers) and another with 11 layers is denoted as DCNN-11 (i.e., four convolutional layers, four max-pooling layers and three fully connected layers) where D means deep.

Each database is shuffled first and processed with data normalization. Then, the databases are split by a ratio of 70/30 into training sets and test sets. To be specific, for ReLD, there are 8086 training samples and 3466 test samples; for ReLD, 28,924 training samples and 12,396 test samples; and for JLD, 37,010 training samples and 15,862 test samples.

The training and test sets of different databases are utilized to train machine learning and deep learning models and evaluate their performance. Results with ReLD, ReLD and JLD are summarized in Tables 3–5, respectively. For each model and database, their accuracy on training and test sets, precision, recall, specificity, F1 score, MCC and OA on test sets are all evaluated. For all three databases, DCNN-11 has the highest OA among all models. The three sets of identification results are projected on the elevation map (Figs. 10a, 11a and 12a) and compared with the corresponding actual landslide distributions (Figs. 10b, 11b and 12b). Statistically, the ReLD database-trained DCNN-11 achieves an area-accuracy of 92.47% in identifying recent landslides. The area-accuracy refers to the percentage of the study area that are correctly identified. The ReLD database-trained DCNN-11 identifies 86.29% area correctly for relict landslides. Also, the JLD database-trained DCNN-11 has a correct landslide identification rate of 88.81%.

Moreover, to evaluate the importance of various predictors, ARPI for all seven predictors are calculated for the three boosting methods (Table 6). For ReLD, the three most important predictors are slope gradient, curvature and TWI. While for ReLD, aspect is the most important predictor, followed by slope gradient and TWI. When it comes to JLD, the three most important predictors are slope gradient, TWI and aspect.

3.5. Model evaluation and analysis results

From the landslide identification results and ranking of the importance of the predictors (Tables 3–6 and Figs. 10–12), some interesting patterns and phenomena can be found:

- (1) Across all three databases, as a result of its strengths in feature extraction and processing multi dimensional data, CNN always has the highest accuracy on the test sets in terms of all accuracy indices (i.e., Accuracy, F1, MCC and OA), followed by the boosting methods, RF and LR. No significant performance differences are observed among the three boosting methods. However, taking the benefits of ensemble learning, all boosting methods lead to satisfactory results. LR, as one of the most widely used machine learning algorithms, has an averaged accuracy above 80%, which is considered quite good as it is the simplest among all the

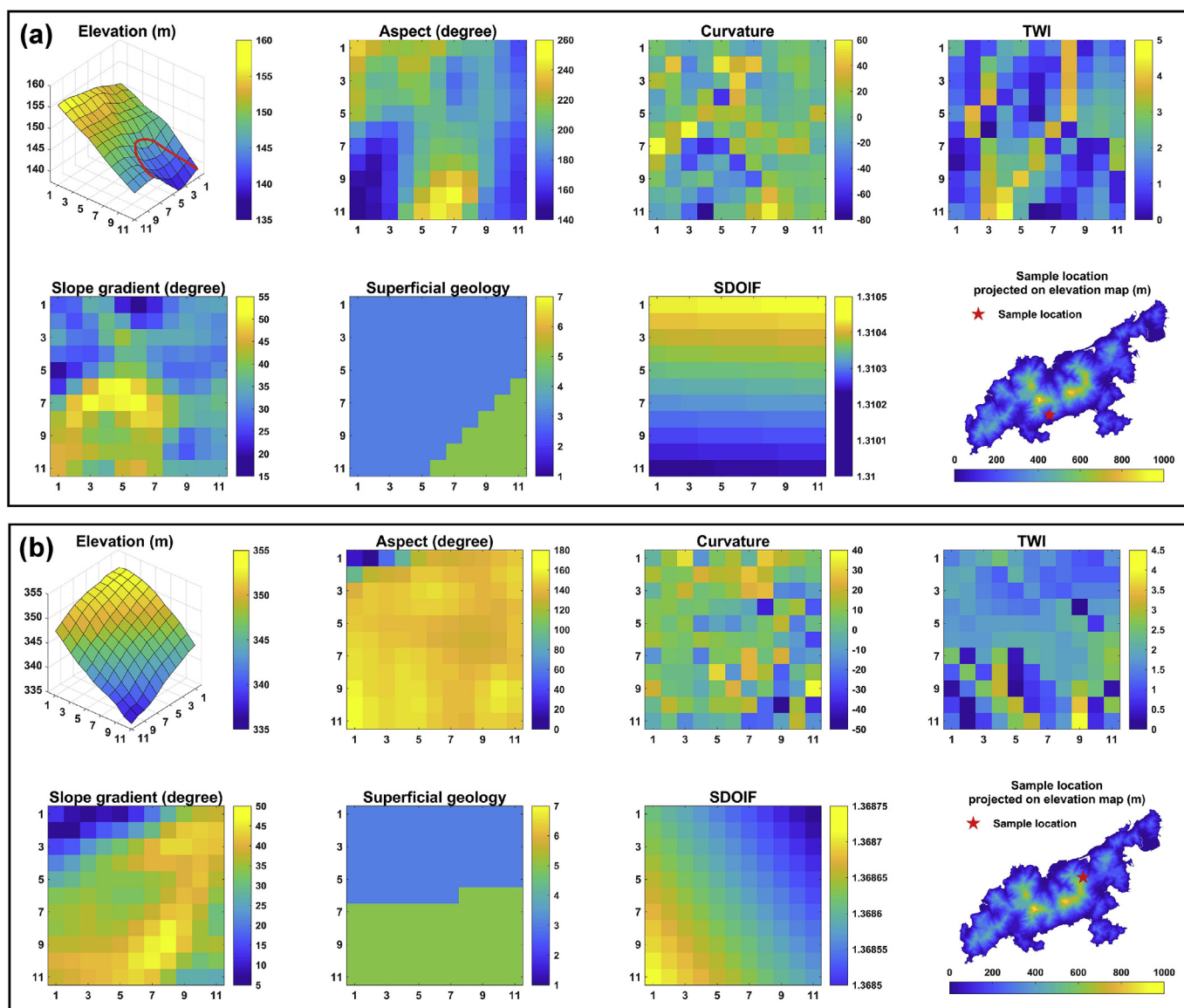


Fig. 8. Examples of (a) a positive sample and (b) a negative sample.

algorithms in this study. The worst performance goes to SVM, whose accuracy is below 80%. As the models are trained with typical configurations in this case study, the ranking of model performance in this case study is not necessarily applicable for other case studies or different model settings. However, the outstanding performance of the proposed landslide identification method has been validated and this case study can be taken as a benchmark for future landslide identification studies.

(2) The difference in the accuracies between the training set and the test set can indicate whether a model is overfitting or not. If the difference is large and the accuracy on the training set is very high, the model is regarded as overfitting. Among the eight machine learning and deep learning models, RF has the highest chance to overfit, followed by the boosting methods. On the contrary, underfitting occurs when a model is not sufficiently sophisticated to precisely capture relationships between the predictors and the response, often leading to low accuracy on both training and test sets. SVM is more susceptible to underfitting than other models, followed by LR. Even though both overfitting and underfitting can be addressed with a variety of methods, this case study will not conduct further tuning as its aim is to provide an overall assessment of the proposed method.

- (3) Among the three databases, the machine learning and deep learning models trained on RecLD have the highest average performance, followed by those trained on JLD. The lowest goes to those trained on ReILD. This result is reasonable as the interpretation result of the relict landslides in ENTLI is not completely accurate and hence may bring certain noises to the databases.
- (4) Among all predictors, slope gradient, aspect, curvature and TWI are the most important according to the results of ARPI.
- (5) The highest area-accuracy on the entire study area goes to the models trained on RecLD, followed by JLD and ReILD. This ranking agrees with the average model performance described earlier. Besides, there is an interesting phenomenon that the area-accuracy is higher than the corresponding DCNN-11 test set accuracy for RecLD and JLD. This is mainly because the specificity of the DCNN-11 models from the two databases is higher than the accuracy (Tables 3 and 5), indicating that these two models are more powerful in identifying non-landslide areas. As the majority of the study area is non-landslide area, the DCNN-11 models trained on RecLD and JLD tend to achieve a higher area-accuracy.
- (6) Taken the Recent Landslide Database (RecLD) as an example, the number of training samples is 8086 that occupy an area of 3.91 km², while the whole Lantau Island covers an area of 147.16 km².

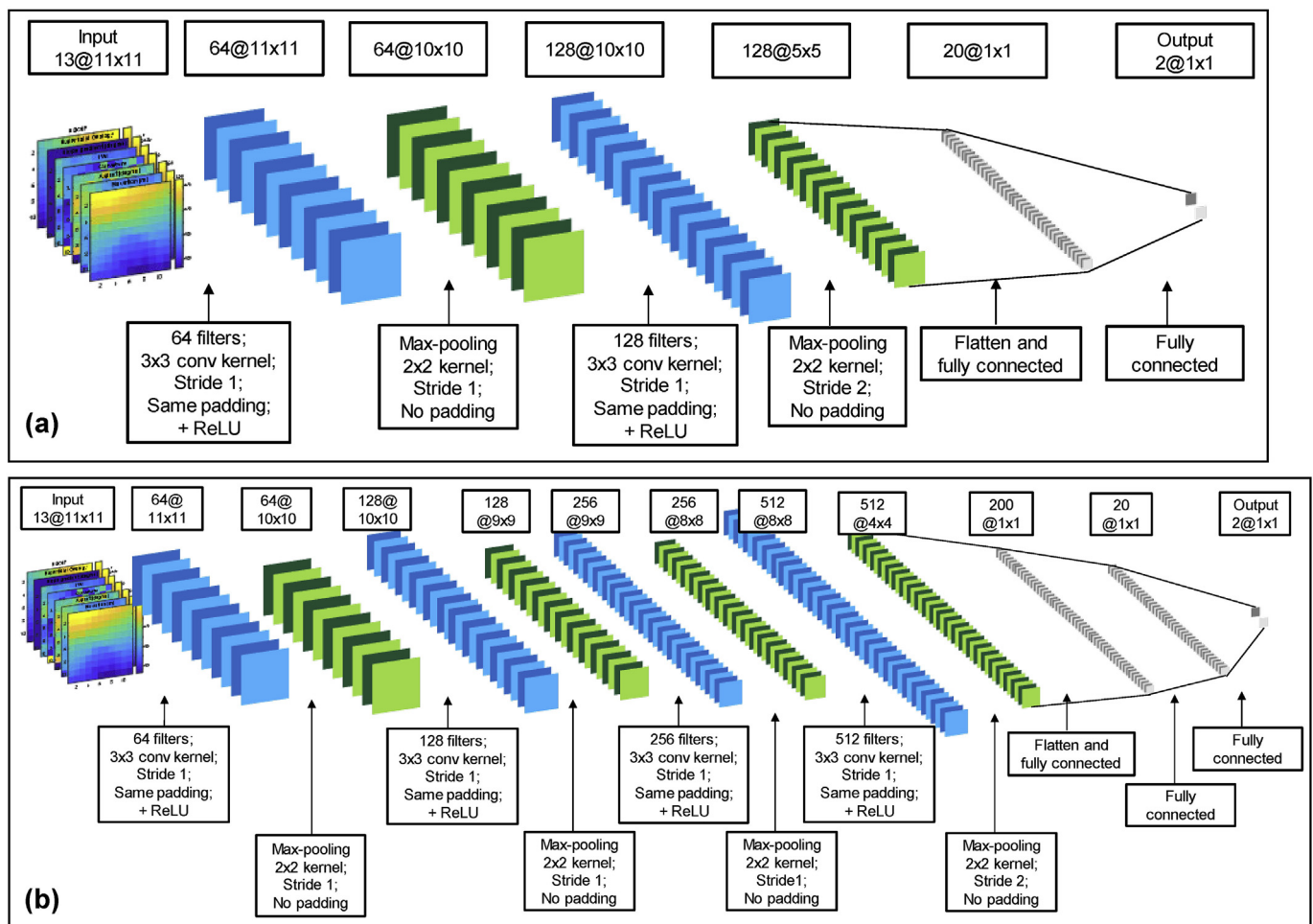


Fig. 9. Two constructed CNN models: (a) CNN-6, a 6-layer CNN model; (b) DCNN-11, an 11-layer deep CNN model.

Table 3

Comparison of the performance of eight machine learning and deep learning models on ReCLD.

Machine learning models	Accuracy on training set	Accuracy on test set	Precision on test set	Recall on test set	Specificity on test set	FPR on rest set	F ₁ score on test set	MCC on test set	OA on test set
LR	0.8492	0.8315	0.8283	0.8283	0.8346	0.1654	0.8283	0.6629	2.3227
SVM	0.7940	0.7726	0.8349	0.6690	0.8725	0.1275	0.7428	0.5542	2.0696
RF	0.9861	0.8396	0.8226	0.8729	0.8050	0.1950	0.8470	0.6801	2.3667
Discrete AdaBoost	0.9175	0.8529	0.8554	0.8554	0.8503	0.1497	0.8554	0.7056	2.4139
LogitBoost	0.9014	0.8592	0.8470	0.8826	0.8350	0.1650	0.8644	0.7188	2.4424
Gentle AdaBoost	0.9217	0.8589	0.8579	0.8661	0.8514	0.1486	0.8620	0.7177	2.4386
CNN-6	0.9168	0.8875	0.8837	0.8968	0.8779	0.1221	0.8902	0.7749	2.5526
DCNN-11	0.9309	0.8932	0.9258	0.8507	0.9343	0.0657	0.8866	0.7886	2.5685

Table 4

Comparison of the performance of eight machine learning and deep learning models on ReLLD.

Machine learning models	Accuracy on training set	Accuracy on test set	Precision on test set	Recall on test set	Specificity on test set	FPR on rest set	F ₁ score on test set	MCC on test set	OA on test set
LR	0.8134	0.7995	0.8027	0.7952	0.8037	0.1963	0.7990	0.5989	2.1974
SVM	0.7505	0.7425	0.7082	0.8269	0.6577	0.3423	0.7630	0.4918	1.9973
RF	0.9906	0.8036	0.8298	0.7663	0.8413	0.1587	0.7968	0.6091	2.2094
Discrete AdaBoost	0.8639	0.8288	0.8368	0.8179	0.8397	0.1603	0.8273	0.6578	2.3139
LogitBoost	0.9133	0.8343	0.8379	0.8298	0.8388	0.1612	0.8339	0.6686	2.3368
Gentle AdaBoost	0.9350	0.8292	0.8344	0.8224	0.8360	0.1640	0.8284	0.6585	2.3161
CNN-6	0.9140	0.8494	0.8761	0.8147	0.8842	0.1158	0.8443	0.7005	2.3942
DCNN-11	0.9026	0.8638	0.8688	0.8579	0.8698	0.1302	0.8633	0.7277	2.4548

Table 5
Comparison of the performance of eight machine learning and deep learning models on JLD.

Machine learning models	Accuracy on training set	Accuracy on test set	Precision on test set	Recall on test set	Specificity on test set	FPR on rest set	F ₁ score on test set	MCC on test set	OA on test set
LR	0.8229	0.8181	0.8197	0.8107	0.8254	0.1746	0.8152	0.6362	2.2695
SVM	0.7295	0.7267	0.7620	0.6509	0.8010	0.1990	0.7021	0.4573	1.8861
RF	0.9899	0.8140	0.8079	0.8307	0.7968	0.2032	0.8191	0.6279	2.2610
Discrete AdaBoost	0.8704	0.8383	0.8323	0.8430	0.8337	0.1663	0.8376	0.6766	2.3526
LogitBoost	0.9092	0.8412	0.8396	0.8416	0.8425	0.1575	0.8406	0.6841	2.3659
Gentle AdaBoost	0.9212	0.8360	0.8326	0.8366	0.8353	0.1647	0.8346	0.6719	2.3425
CNN-6	0.8995	0.8725	0.8759	0.8649	0.8800	0.1200	0.8704	0.7450	2.4879
DCNN-11	0.9012	0.8750	0.8809	0.8642	0.8856	0.1144	0.8725	0.7500	2.4975

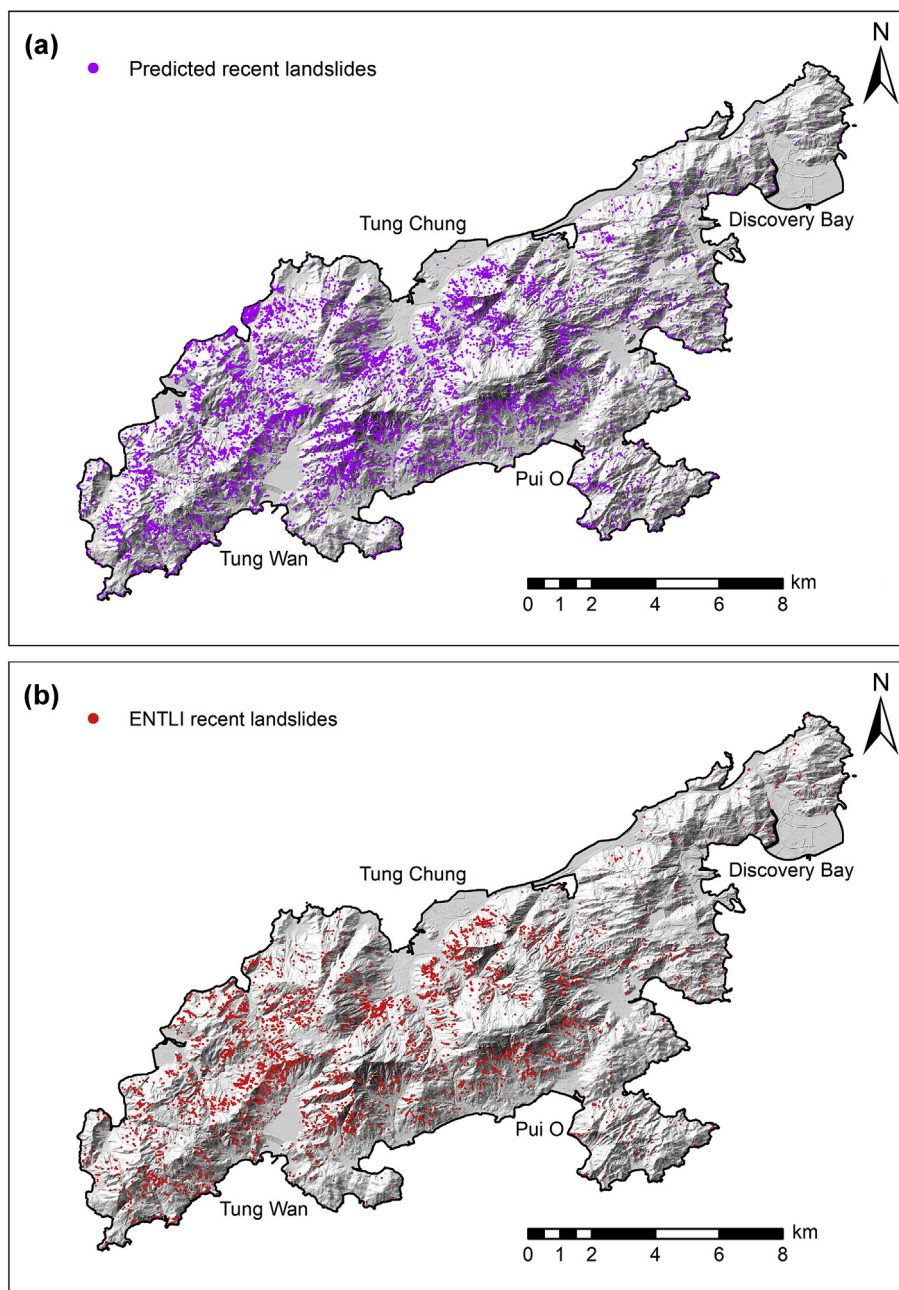


Fig. 10. ReclD based machine learning results: (a) predicted recent landslides using DCNN-11 and (b) recent landslides in ENTLLI.

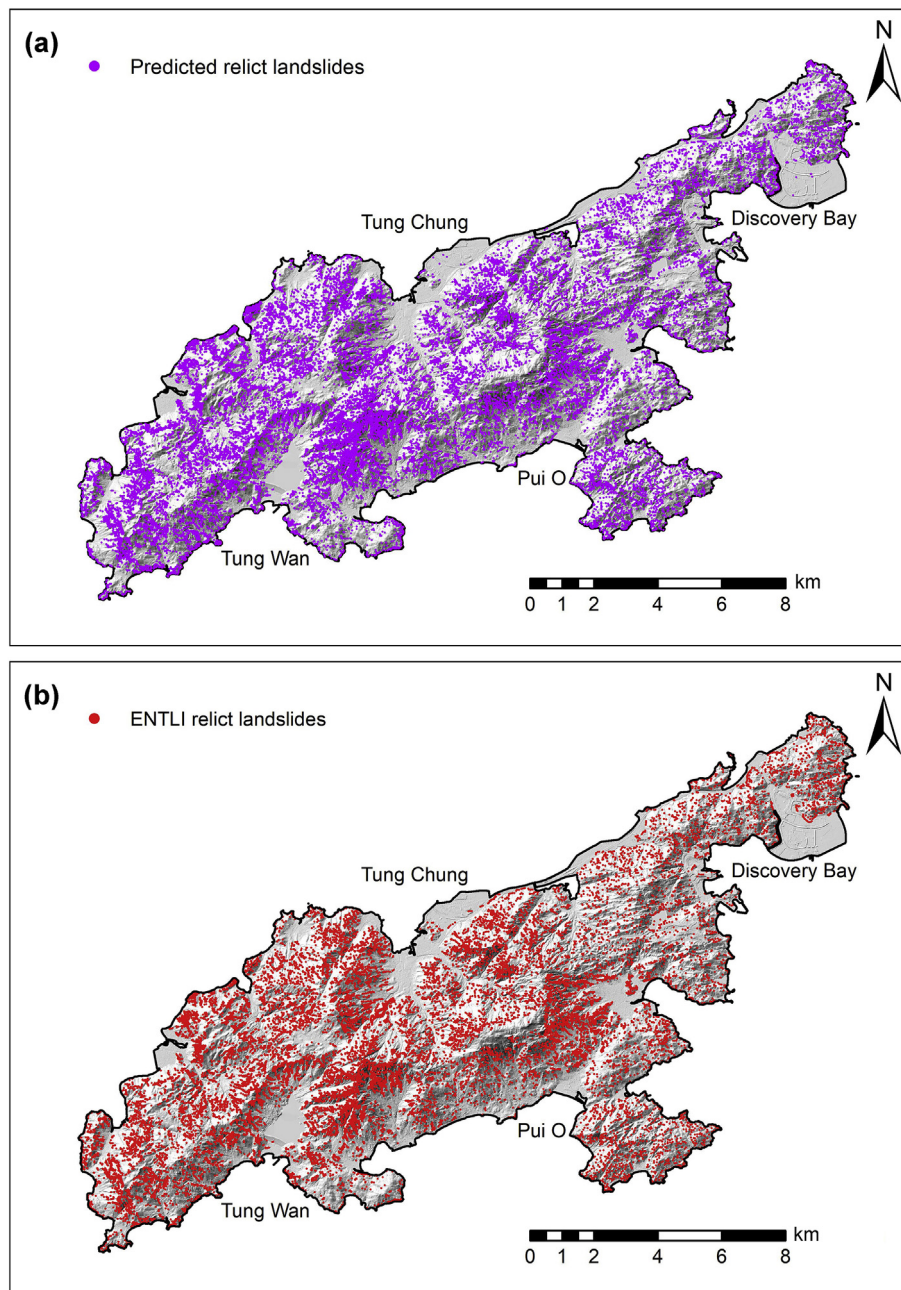


Fig. 11. RelLD based machine learning results: (a) predicted relict landslides using DCNN-11 and (b) relict landslides in ENTLLI.

Only the information of 2.66% area is used during model training while the proposed method is able to give an overall area-accuracy of 92.47%. In the light of this, the general applicability and the adaptability of the proposed method have been proven promising in the case study.

4. Limitations

Although the proposed method works well in the case study, some limitations should be pointed out. First, the ENTLLI covers the landslides ranging from 1924 to 2009 while the DTM utilized was captured in 2011, which implies some inconsistencies in the terrain data and the landslide data. This may lead to underestimated performance of the proposed method. Second, as described before, the relict landslide records of ENTLLI are not completely accurate. Third, CNNs with more layers should be investigated in the future when higher computational power is available.

5. Conclusions

This paper proposes an integrated method for landslide identification using machine learning and deep learning. A case study on Lantau Island, Hong Kong, with multiple landslide databases, is conducted to illustrate and validate the proposed method. The following conclusions can be drawn:

- (1) Among all eight machine learning and deep learning models (i.e., LR, SVM, RF, Discrete AdaBoost, LogitBoost, Gentle Adaboost, CNN-6 and DCNN-11), DCNN-11 is found to be the most promising model to tackle the landslide identification problem.
- (2) Among the three landslide databases (i.e., RecLD, RelLD and JLD), models trained on RecLD show the highest averaged identification accuracy, 89.3% with the DCNN-11 model. The highest accuracies trained on JLD and RelLD are 87.5% and 86.4%, respectively.

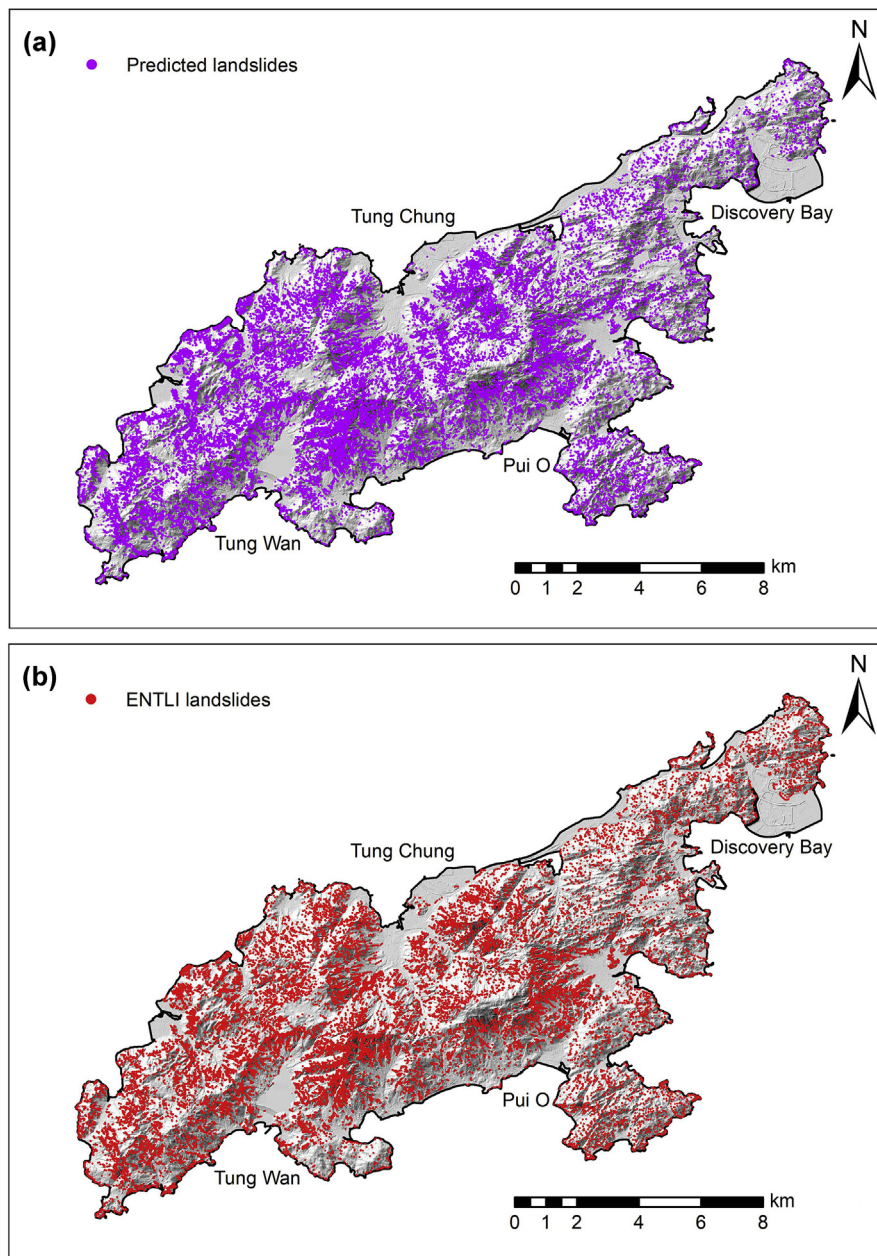


Fig. 12. JLD based machine learning results: (a) predicted landslides using DCNN-11 and (b) landslides in ENTLI.

Table 6
Results of ARPI based on models trained with three Boosting methods.

Database	Elevation (m)	Aspect (°)	Curvature	Slope gradient (°)	TWI	Superficial geology	SDOIF
RecLD	0.0358	0.0459	0.2121	0.5089	0.1320	0.0421	0.0232
RelLD	0.0128	0.4068	0.0669	0.3727	0.1138	0.0175	0.0095
JLD	0.0184	0.0827	0.0806	0.6524	0.1215	0.0327	0.0118

(3) Slope gradient, aspect, curvature and TWI are shown to be more important than the other three predictors (i.e., elevation, superficial geology and SDOIF).

(4) The area identification accuracy of DCNN-11 is as high as 92.5% when trained with RecLD, 88.6% when trained with JLD, and 86.3% with RelLD. These show remarkable performance and robustness of the proposed landslide identification method.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by the Research Grants Council of the Hong Kong SAR Government (Nos. 16205719, AoE/E-603/18 and 16206217). The authors would also like to thank the Geotechnical Engineering Office of Civil Engineering and Development Department for providing the DTM and landslide inventory data.

References

- AECOM, Lin, B., 2015. 24-hour Probable Maximum Precipitation Updating Study. GEO Report No. 314. Geotechnical Engineering Office, Hong Kong Special Administration Region.
- Barlow, J., Martin, Y., Franklin, S.E., 2003. Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Can. J. Rem. Sens.* 29 (4), 510–517.
- Chen, H.X., Zhang, S., Peng, M., Zhang, L.M., 2016. A physically-based multi-hazard risk assessment platform for regional rainfall-induced slope failures and debris flows. *Eng. Geol.* 203, 15–29.
- Ching, J., Phoon, K.K., 2018. Constructing site-specific multivariate probability distribution model using Bayesian machine learning. *J. Eng. Mech.* 145 (1), 04018126.
- Ding, A., Zhang, Q., Zhou, X., Dai, B., 2016. Automatic recognition of landslide based on CNN and texture change detection. In: Proceedings of the Chinese Association of Automation (YAC), Youth Academic Annual Conference, Wuhan, China, 11–13 November 2016. IEEE, pp. 444–448.
- Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. *Ann. Stat.* 28 (2), 337–407.
- Friedman, J., Hastie, T., Tibshirani, R., 2001. *The Elements of Statistical Learning*. Springer Series in Statistics, New York.
- Gao, L., Zhang, L.M., Lu, M., 2017. Characterizing the spatial variations and correlations of large rainstorms for landslide study. *Hydrol. Earth Syst. Sci.* 21 (9), 4573.
- Gao, L., Zhang, L.M., Cheung, R.W.M., 2018. Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. *Landslides* 15, 727–740.
- Geo, 1996. *Compilation of a Database on Landslide Consequence*. Submitted by Mitchell, McFarlane, Brentnall & Partners Int. Ltd., Agreement No. GEO 8/95. Geotechnical Engineering Office, Hong Kong Special Administration Region.
- Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Rem. Sens.* 11 (2), 196.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.
- Griffiths, D.V., Huang, J., Fenton, G.A., 2011. Probabilistic infinite slope analysis. *Comput. Geotech.* 38 (4), 577–584.
- Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* 112 (1–2), 42–66.
- Haneberg, W.C., Cole, W.F., Kasali, G., 2009. High-resolution lidar-based landslide hazard mapping and modeling, UCSF Parnassus Campus, San Francisco, USA. *Bull. Eng. Geol. Environ.* 68 (2), 263–276.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*. Springer, New York.
- Keypour, R.N., Oommen, T., Martha, T.R., Sajinkumar, K.S., Gierke, J.S., 2018. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinf.* 64, 1–11.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.* 7 (1), 3–10.
- Li, X.J., Cheng, X.W., Chen, W.T., Chen, G., Liu, S.W., 2015. Identification of forested landslides using Lidar data, object-based image analysis, and machine learning algorithms. *Rem. Sens.* 7 (8), 9705–9726.
- Li, X.Y., Zhang, L.M., Zhang, S., 2018. Efficient Bayesian networks for slope safety evaluation with large quantity monitoring information. *Geosci. Front.* 9 (6), 1679–1687.
- Li, X.Y., Zhang, L.M., Xiao, T., Zhang, S., Chen, C., 2019. Learning failure modes of soil slopes using monitoring data. *Probabilist. Eng. Mech.* 56, 50–57.
- Lo, M.K., Leung, A.Y., 2019. Bayesian updating of subsurface spatial variability for improved prediction of braced excavation response. *Can. Geotech. J.* 56 (8), 1169–1183. <https://doi.org/10.1139/cgj-2018-0409>.
- Martha, T.R., Kerle, N., Jetten, V., van Westen, C.J., Kumar, K.V., 2010. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology* 116 (1–2), 24–36.
- Martha, T.R., Kerle, N., Van Westen, C.J., Jetten, V., Kumar, K.V., 2012. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. *ISPRS J. Photogrammetry Remote Sens.* 67, 105–119.
- Maunsell-Fugro Joint Venture, Geo, 2007. Final Report on Compilation of the Enhanced Natural Terrain Landslide Inventory (ENTLI), Maunsell-Fugro Joint Venture & Geotechnical Engineering Office. Hong Kong Special Administration Region.
- McKean, J., Roering, J., 2004. Objective landslide detection and surface morphology mapping using high-resolution airborne laser altimetry. *Geomorphology* 57 (3–4), 331–351.
- Moosavi, V., Talebi, A., Shirmohammadi, B., 2014. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology* 204, 646–656.
- Naidu, S., Sajinkumar, K.S., Oommen, T., Anuja, V.J., Samuel, R.A., Muraliedharan, C., 2018. Early warning system for shallow landslides using rainfall threshold and slope stability analysis. *Geosci. Front.* 9 (6), 1871–1882.
- Papaioannou, I., Straub, D., 2017. Learning soil parameters and updating geotechnical reliability estimates under spatial variability—theory and application to shallow foundations. *Georisk* 11 (1), 116–128.
- Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* 180, 60–91.
- Van Den Eeckhaut, M., Kerle, N., Poesen, J., Hervás, J., 2012. Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data. *Geomorphology* 173, 30–42.
- Wang, H.J., Xiao, T., Li, X.Y., Zhang, L.L., Zhang, L.M., 2019. A novel physically-based model for updating landslide susceptibility. *Eng. Geol.* 251, 71–80.
- Xu, C., 2015. Preparation of earthquake-triggered landslide inventory maps using remote sensing and GIS technologies: principles and case studies. *Geosci. Front.* 6 (6), 825–836.
- Zhang, S., Zhang, L.M., Peng, M., Zhang, L.L., Zhao, H.F., Chen, H.X., 2012. Assessment of risks of loose landslide deposits formed by the 2008 Wenchuan earthquake. *Nat. Hazards Earth Syst. Sci.* 12 (5), 1381–1392.
- Zhang, W., Goh, A.T., Zhang, Y., Chen, Y., Xiao, Y., 2015. Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. *Eng. Geol.* 188, 29–37.
- Zhang, W., Goh, A.T., 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* 7 (1), 45–52.