Incorporating mitigation strategies in machine learning for landslide susceptibility prediction

Hai-Min Lyu, Zhen-Yu Yin, Pierre-Yves Hicher, Farid Laouafa

Geoscience Frontiers 15 (2024) 101869

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research Paper

Incorporating mitigation strategies in machine learning for landslide susceptibility prediction

Hai-Min Lyu [a], Zhen-Yu Yin [b,*], Pierre-Yves Hicher [c], Farid Laouafa [d]

[a] Key Laboratory for Resilient Infrastructures of Coastal Cities (MOE), College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China
[b] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[c] Research Institute of Civil Engineering and Mechanics (GeM), UMR CNRS 6183, Ecole Centrale de Nantes, France
[d] National Institute for Industrial Environment and Risks (INERIS), Verneuil-en-Halatte, France

ARTICLE INFO

Article history:
Received 5 September 2023
Revised 10 April 2024
Accepted 16 May 2024
Available online 18 May 2024
Handling Editor: B. Pradhan

Keywords:
Machine learning
Landslide susceptibility
Spatial prediction
Mitigation strategies

ABSTRACT

This study proposes an approach that considers mitigation strategies in predicting landslide susceptibility through machine learning (ML) and geographic information system (GIS) techniques. ML models, such as random forest (RF), logistic regression (LR), and support vector classification (SVC) are incorporated into GIS to predict landslide susceptibilities in Hong Kong. To consider the effect of mitigation strategies on landslide susceptibility, non-landslide samples were produced in the upgraded area and added to randomly created samples to serve as ML models in training datasets. Two scenarios were created to compare and demonstrate the efficiency of the proposed approach; Scenario I does not considering landslide control while Scenario II considers mitigation strategies for landslide control. The largest landslide susceptibilities are 0.967 (from RF), followed by 0.936 (from LR) and 0.902 (from SVC) in Scenario II; in Scenario I, they are 0.986 (from RF), 0.955 (from LR) and 0.947 (from SVC). This proves that the ML models considering mitigation strategies can decrease the current landslide susceptibilities. The comparison between the different ML models shows that RF performed better than LR and SVC, and provides the best prediction of the spatial distribution of landslide susceptibilities.

1. Introduction

Landslides are among the most common natural hazards that contribute to widespread loss of life and property each year (Liao et al., 2022; Abraham et al., 2023). There were 378 major landslides worldwide between 1998 and 2017 that killed 18,414 people and caused economic losses of up to $8 billion (Liao et al., 2022). Due to its specific topography, Hong Kong is frequently affected by landslides (Cheung, 2021; Ng et al., 2021a, 2021b); previous reports list over 300 landslides that occur annually, most of which are induced by rainstorms. Hong Kong has a subtropical climate with an annual average rainfall (AAR) of 2400 mm (Cheung, 2021). According to the Hong Kong Observatory (HKO), annual rainfall has been increasing at a mean rate of 26 mm per decade due to climate change (Cheung, 2021; Ju et al., 2022; Xiao et al., 2022). The increased frequency and severity of rainstorms caused by climate change have worsened landslide events. Clusters of urban developments in addition to the high population concentration and other vulnerable facilities suggest that landslides in Hong Kong's urban areas could have serious consequences.

A detailed map of landslide susceptibility is efficient in delineating areas with the most risk to landslides (Dematteis et al., 2022; Feng et al., 2022). Predicting the susceptibility of a landslide in specific anthropogenic and environmental situations is the highest degree of development in terms of a planning tool for controlling landslides and mitigating their consequences (Lombardo et al., 2021; Park and Lee, 2022). Mapping landslide susceptibility is another approach for risk mitigation, and published data suggests that it is a very popular topic in studies on the prediction of landslide occurrence and landslide management decisions (Wubalem, 2021).

Geographic information systems or GIS have demonstrated a remarkable capacity for spatial modelling and analysis of landslides, as well as for susceptibility mapping (Ng et al., 2021a, 2021b; Wang et al., 2022). Many GIS-technology-based approaches that are conducted rely on the complexity of the models and specificities of the study area; existing approaches can be divided into physics-based, statistical, and data-driven methods (Li et al., 2022; Shen et al., 2023; Lyu et al., 2024). Physics-based methods consider the physical mechanisms that lead to slope failure. How-

* Corresponding author.
E-mail address: zhenyu.yin@polyu.edu.hk (Z.-Y. Yin).

https://doi.org/10.1016/j.gsf.2024.101869
1674-9871/

ever, these methods require accurate information regarding the mechanical properties of the soil constituting the slope. Physics-based methods tend to be more accurate in describing the development of a specific landslide but are less suitable for regional landslide evaluations. The GIS-based multi-criteria decision-making (MCDM) is a typical statistical method used to analyse the relationship between different driving factors and to quantitatively assess the probability of landslide occurrence in a study area (Salehpour Jam et al., 2021; Tyagi et al., 2021). Among the MCDM methods, the analytical hierarchy process (AHP), fuzzy AHP, analytical network process (ANP), and fuzzy logic have been used to calculate the relative importance of the factors that influence landslides the most. MCDM methods are primarily based on pair-wise comparisons of expert opinions. Data-driven approaches, specifically the machine learning (ML) methods, constitute a recent popular approach to solve spatial landslide modelling problems owing to their simplicity and low cost (Liu et al., 2020).

Compared with statistical methods such as MCDM, hybrid ML-GIS methods usually provide better performance in generating landslide susceptibility maps as well as a higher data processing speed. Moreover, ML methods have proven their efficiency in landslide susceptibility evaluation as evidenced by improved global attention (Park et al., 2021; Khezri et al., 2022; Wang et al., 2022). Many ML methods have been applied to probability analysis, including ensemble methods, such as random forest (RF) and adaboost tree (Sun et al., 2020); to linear methods, such as logistic regression (LR), support vector machine (which is classified into support vector regression (SVR) and support vector classification (SVC)); and to neural methods, such as artificial neural networks (ANNs) and convolutional neural networks (CNNs) (Huang et al., 2017; Hong et al., 2018; Wang et al., 2020; Ullah et al., 2022). Pradhan et al. (2021, 2023a, 2023b) proposed a series of explainable ML models to predict spatial flood susceptibility. The explainable ML models provide interpretations of predictive results from artificial intelligence models (Al-Najjar et al., 2023; Abraham et al., 2023). Yi et al. (2020) proposed the SHAP-XGB model to predict landslide susceptibility taking into account the local geospatial heterogeneity.

Previous studies suggest that it is difficult to select a perfect algorithm in ML modelling for landslide susceptibility prediction (Sun et al., 2021; Ng et al., 2021a, 2021b). A range of parameters, such as data size, reliability, and computation time are considered. Except for tree-based models (e.g. RF), the performance of linear models (e.g. LR, SVR, and SVC) is generally contingent to a standardisation or normalisation process. Moreover, hyperparameters are tuned to achieve the optimum predictions (Wang et al., 2021a; Ma et al., 2022; Lyu and Yin, 2023). Furthermore, performance metrics are often used to evaluate the performance of various algorithms.

GIS possesses a powerful database function that can be integrated with ML models to produce a robust tool for predicting landslide susceptibility. A few studies have been conducted in Hong Kong (Ng et al., 2021a, 2021b; Xiao et al., 2022) but they do not consider the influence of mitigation strategies on landslide management nor do they provide a clear description on how to incorporate ML methods into GIS to estimate landslide susceptibilities. This is a critical aspect for landslide susceptibility prediction. Landslide susceptibility is continuously affected by present-day mitigation measures. Thus, during the prediction of landslide susceptibility, it is also important to consider the influence mitigation measures might have on landslide management. These measures are a determining factor in abating landslide occurrence at the regional level, in addition to helping reduce the risks of fatalities and reducing economic damage. However, their seamless integration in assessment models, particularly ML models, remains burdensome. This necessitates the development of comprehensive and robust models for tackling this issue and improving risk management practices.

This study proposes an approach to consider and evaluate the effect of mitigation strategies on landslide susceptibility in ML models. The objectives of this study are: (i) to develop an approach to consider and evaluate mitigation strategies on landslide susceptibility; (ii) to predict landslide susceptibilities by producing non-landslide samples in upgraded areas to generate a dataset for the ML training process; (iii) deduce the best ML models to incorporate into GIS for predicting landslide susceptibilities considering mitigation strategies. The major contribution of this study, that sets it apart from previous research (Ng et al., 2021a, 2021b; Cheung, 2021; Xiao et al., 2022), lies in the innovative evaluation of landslide susceptibilities by considering the effects of mitigation strategies in ML. This study proves that mitigation strategies can substantially decrease current landslide susceptibility—a pivotal aspect not considered in previous studies. The novelty of this study can be summarized as follows: (i) landslide susceptibilities are predicted by considering the mitigation strategies on landslide management; (ii) the effects of mitigation strategies are evaluated by producing non-landslide samples in upgraded areas during the ML training process; (iii) mechanical properties of surface soils are considered in predicting landslide susceptibilities.

2. Study area and database

2.1. Study area

Hong Kong is located between 22°10′N–22°30′N and 113°50′E–114°20′E; it encompasses a land area of 1106.7 km^2 that includes the Hong Kong Island, Kowloon Peninsula, and the New Territories. Owing to its unique geology, topography, and the receipt of heavy rainfall, landslides are a frequent occurrence in Hong Kong. Fig. 1 shows the spatial distribution of landslides from 1985 to 2021. According to statistics from a Geotechnical Engineering Office report (GEO, 2000), there were 9843 historical landslide events from 1985 to 2021, which can be classified into five groups according to the displaced volumes: very minor (<5 m^3), minor (5–50 m^3), medium (50–100 m^3), major (100–500 m^3), and very major (>500 m^3). Based on the spatial distribution of these historical landslide events, the 9843 landslide pixels dataset also constitute the basis for establishing a ML model to predict landslide susceptibility (GEO, 2000). In addition to landslide records, the database in ML models includes rainfall, topographical, geological, and anthropogenic conditions.

2.2. Rainfall

Rainfall is among the primary factors that prompt landslides. To analyse the relationship between landslides and rainfall, the AAR and various occurrences of landslides between 1985 and 2021 are collected and compared (Fig. 2). The number of landslides increase with an increase in the AAR; for example, the June 6–9, 2008 rainstorm caused a sizeable number of landslides throughout the region (Lam et al., 2012). In addition to the amount of rainfall, its spatial distribution is also critical to the location of landslides. Based on collected reports of rainfall amounts and rain gauge locations, the spatial distribution of the AAR can be obtained using GIS tools. The spatial distribution of the AAR from 2011 to 2021 shows that Hong Kong Island, the Kowloon Peninsula, and the southern part of the New Territories recorded between 1732 to 2565 mm of ARR (Fig. 3). Information on rainfall data and rain gauges were obtained from the Hong Kong Observatory (HKO) and GEO. The data sources and descriptions of the rainfall factors are listed in Table 1.

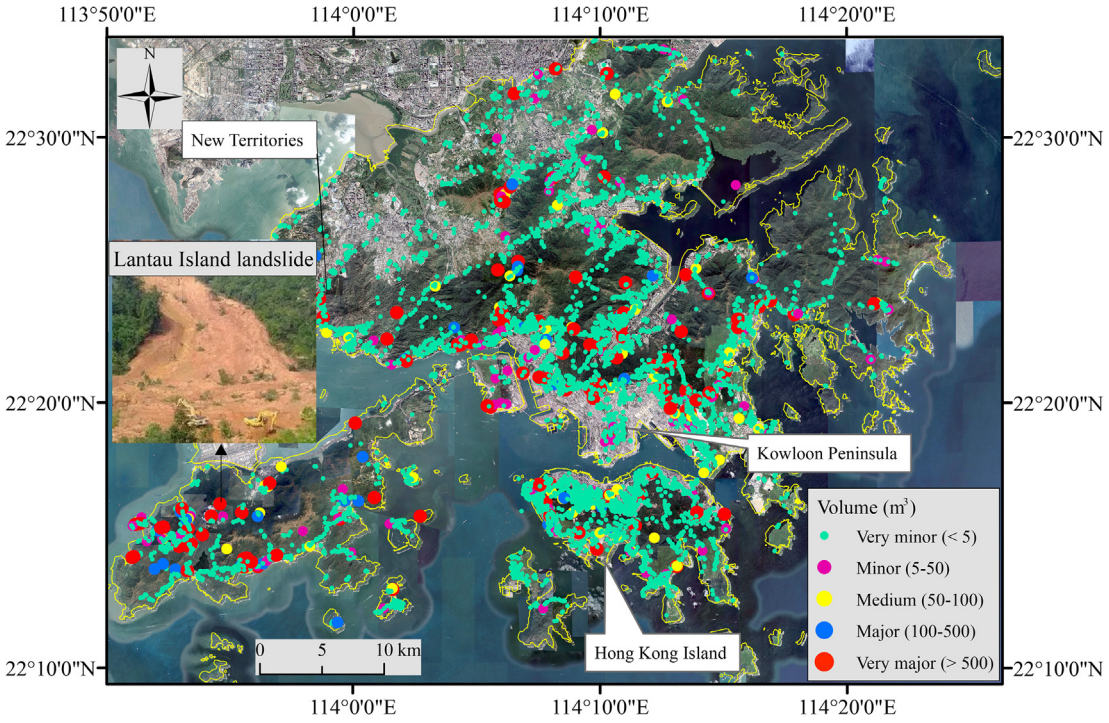

Fig. 1. Spatial distribution of landslides between 1985 and 2021.

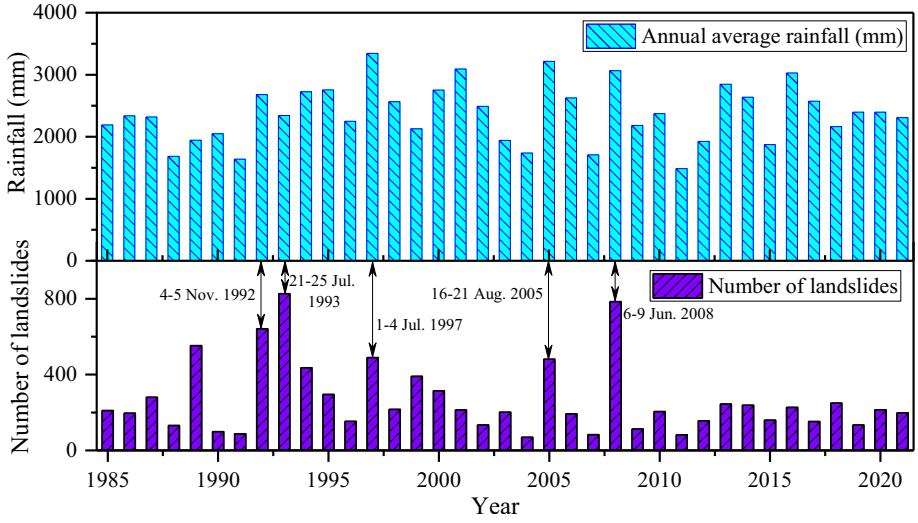

Fig. 2. Recorded AAR and number of landslides between 1985 and 2021.

2.3. Topographic conditions

Elevation and slope are typical factors that reflect the topographical characteristics of the study area. The curvature refers to the direction of water flow on the surface. The profile and plane curvatures are parallel and perpendicular to the direction of the maximum slope, respectively, which helps understand the flow of water across a surface. Moreover, this aspect is used to define the compass direction of the slope face, which reflects vegetation and moisture retention. The topographic wetness index (TWI) is a measure of surface runoff and water aggregation in a basin, which can be calculated using Eq. (1),

$$\mathrm{TWI} = \ln(A/\tan B) \quad (1)$$

where A is the area of the upstream catchment and B is the slope angle in radians. In this study, a digital terrain model (DTM) with 5 m resolution provided by the Survey and Mapping Office of the Lands Department (SMLD) of Hong Kong was used to extract topographic characteristics. Using this DTM, maps of the elevation (Fig. 4a), slope (Fig. 4b), plane curvature (Fig. 4c), profile curvature (Fig. 4d), aspect (Fig. 4e), and TWI (Fig. 4f) were produced. To consider the influence of road networks and the river system, densities of the road network (Fig. 4g) and the river system (Fig. 4h) were obtained using GIS tools. The data sources and descriptions of the topographic conditions are presented in Table 1.

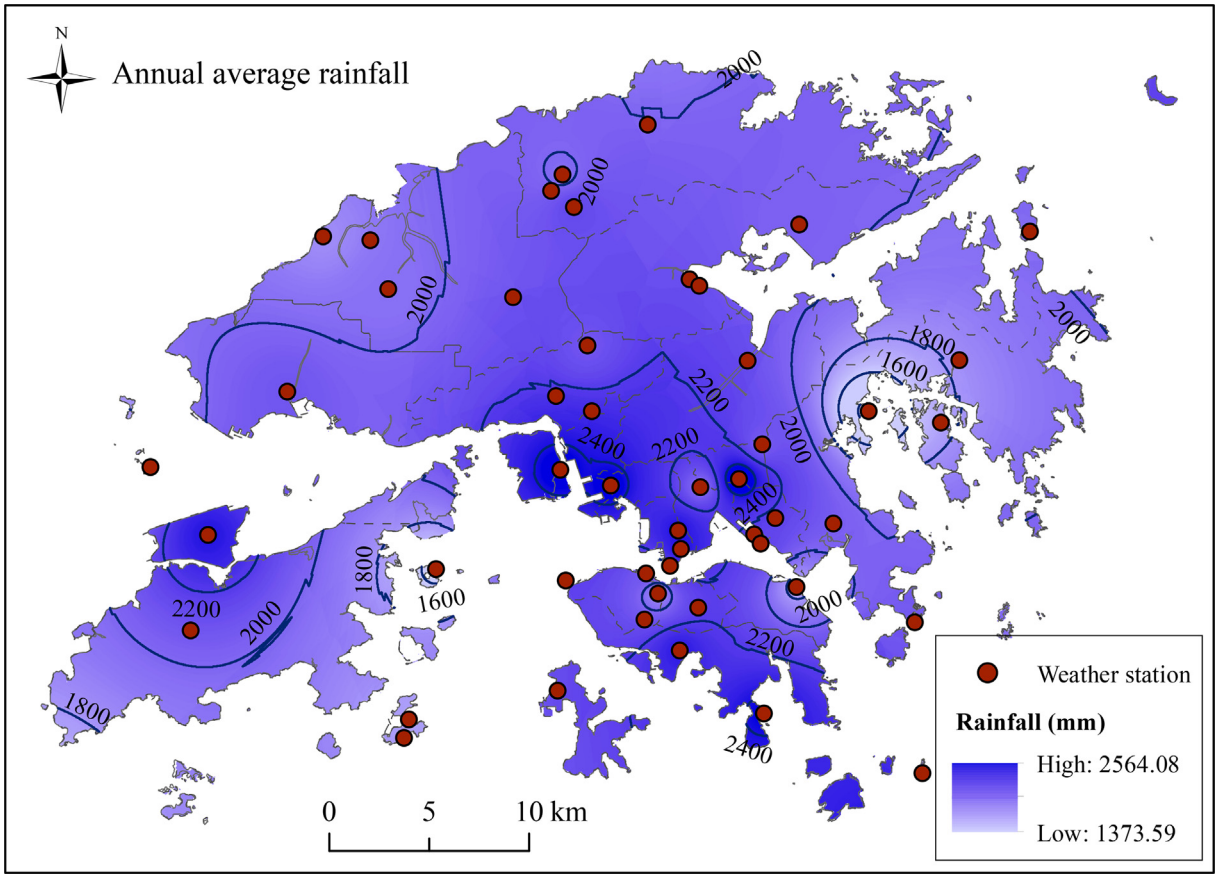

Fig. 3. Spatial distribution of AAR between 2011 and 2021.

Table 1
Descriptive statistics and sources of the databases considered.

Factor	Description	Type	Statistics				Data source/resolution/type
			Min	Max	Mean	STD	
Average rainfall (mm)	Annual average rainfall (2010–2020)	N, S	0	2563.89	2058.04	183.64	Rain gauges and annual rainfall reports from GEO/10 m/raster
Elevation (m)	Digital elevation of terrain surface	N, S	0	925	108.71	124.30	DTM/5 m/raster
Slope (°)	Angle of slope inclination	N, S	0	49.04	11.57	8.90	
Plane curvature	Curvature perpendicular to the slope, indicating concave or convex surfaces	N, S	−1.10	1.48	0.02	0.19	
Profile curvature	Curvature parallel to the slope, indicating concave or convex surfaces	N, S	−1.84	2.41	0.03	0.21	
Aspect	Compass direction of slope exposure	N, S	0	254	167.46	45.13	
TWI	Topographic wetness index, measures water aggregation and surface runoff	N, S	−0.03	23.96	6.97	3.98	
Road network density	Spatial distribution of road network density	N, S	0	3391	1876	65.34	SMLD/10 m/raster
River system density	Spatial distribution of river system density	N, S	0	4.86	2.64	2.15	DTM/10 m/raster
Geology	Geological conditions in study area	C, S	/	/	/	/	GEO-map/1:100,000/polygon
Land use	Land cover of study area	C, S	/	/	/	/	SMLD/-/Polygon
Soil mechanics c	Soil cohesion	N, S	0	15	9.87	3.99	GEO-map, Geoguide 1/-/Polygon
f (°)	Soil internal friction angle	N, S	0	44	35.71	8.44	GEO-map, Geoguide 1/-/Polygon
Infrastructure	Distance to buildings and metro lines	N, S	0	500	126.16	170.60	SMLD/-/Polygon + polyline + point

Note: N, C, and S stand for numerical, categorical, and static variables, respectively. SMLD stands for Survey and Mapping Office of the Land Department.

2.4. Geologic and anthropogenic conditions

Maps of the geographic and anthropogenic factors controlling landslides are discussed in this section (Fig. 5). A geology map (Fig. 5a) with 1:100,000 obtained from GEO was adopted to derive the geological conditions of the region. The geological map was firstly converted into a raster image to extract the geological conditions at the locations of the landslide pixels. To consider the influence of faults, contours marking the areas with proximities of 100, 200, 300, 400, and 500 m to the faults were produced using

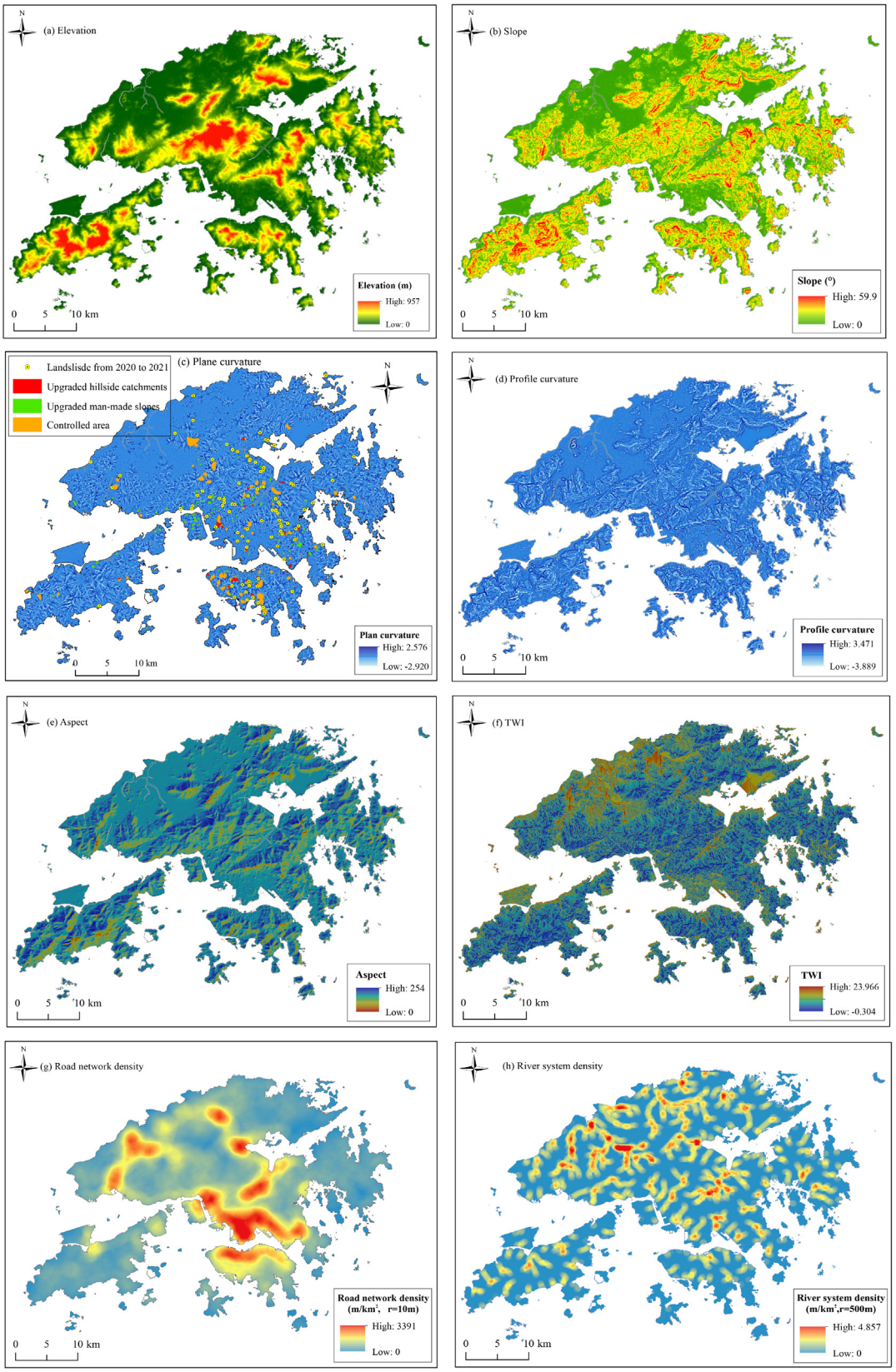

Fig. 4. Spatial distribution of topographic variables: (a) elevation; (b) slope; (c) plane curvature; (d) profile curvature; (e) aspect; (f) TWI.

GIS tools. The geology was classified into eight levels according to age data. Table 2 provides a detailed description of the geology. This study innovatively integrates the mechanical properties of surface soil to the regional landslide susceptibility prediction, in contrast with previous studies that simply adopted regional geological maps to reflect the geological conditions (Ng et al., 2021a,

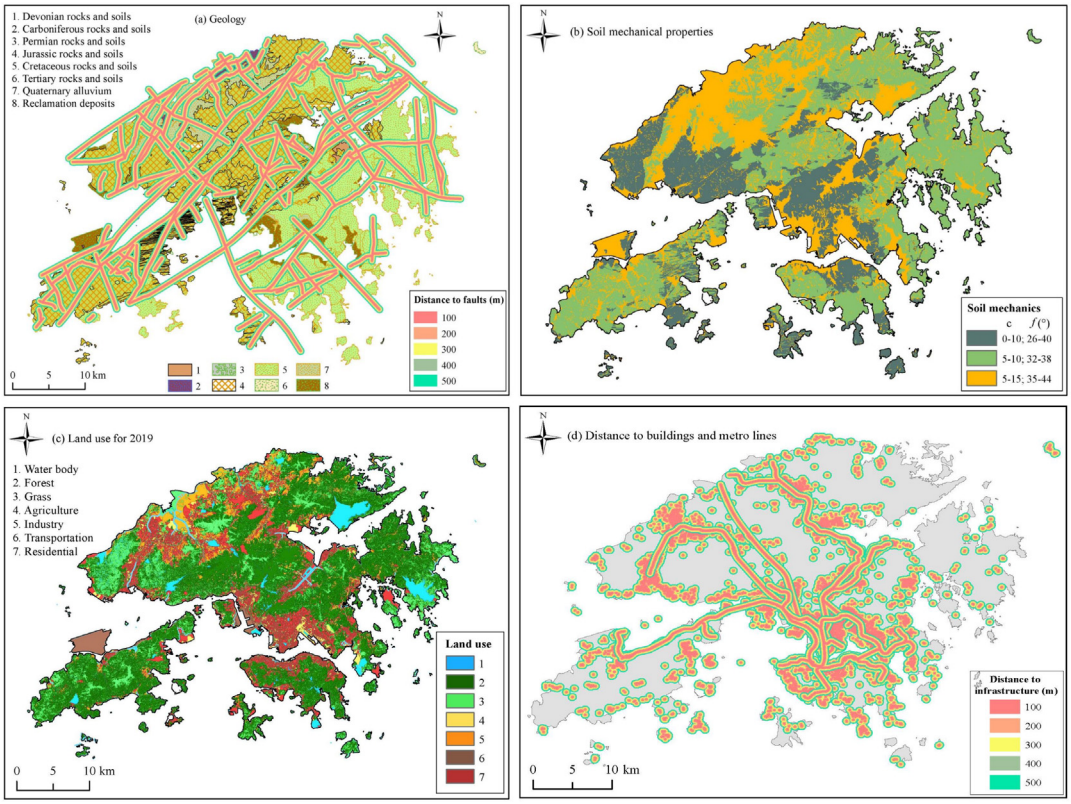

Fig. 5. Geographic and anthropogenic factors: (a) geology; (b) soil mechanical properties; (c) land use; (d) distance to infrastructure.

Table 2
Description of geology.

Level	Category	Area (%)	Landslide pixel (%)	Description
1	Devonian rocks and soils	1.01	0.03	Devonian quartz sandstone and siltstone
2	Carboniferous rocks and soils	1.32	1.06	Carboniferous metamorphosed sandstone and carbonaceous siltstone
3	Permian rocks and soils	0.19	0.29	Permian calcareous sandstone
4	Jurassic rocks and soils	49.01	47.40	Jurassic sandstone, siltstone and mudstone
5	Cretaceous rocks and soils	31.13	47.89	Cretaceous vitric tuff and rhyolite lava
6	Tertiary rocks and soils	0.10	0	Tertiary calcareous siltstone with rare chert interbeds
7	Quaternary alluvium	10.89	2.51	Quaternary alluvium on valley floors and colluvium on valley sides
8	Reclamation deposits	4.36	0.83	Reclamation deposits dominantly composed of marine sand and rock

2021b; Xiao et al., 2022). In this study, both, the mechanical properties of the soil and geological maps of Hong Kong were used in the ML models. Based on the surface soil type and following previously established guidelines, the cohesion (c) and friction angle (f) were classified into three levels (GEO, 2000; Fig. 5b) Table 3 lists the description of mechanical properties of surface soils. In addition to geological conditions, anthropogenic conditions such as land use, buildings, and infrastructure were also considered. The 2019 land use data is classified into seven types (Fig. 5c). Table 4 provides detailed information regarding land use, which were converted into raster data to extract relevant information at the location of the landslide pixels. To consider the importance of infrastructure, contours marking the proximity of buildings and metro lines were produced in GIS using a buffer analysis with distances of 100, 200, 300, 400, and 500 m. The risk of landslides and their damaging effects increase as the distance to infrastructure, buildings, and metro lines decreases. The data sources and descriptive statistics for the geological and anthropogenic conditions are listed in Table 1.

2.5. Strategies for landslide control

A series of appropriate mitigation strategies to manage landslide susceptibility are adopted by the government each year.

Table 3
Description of geotechnical parameters for surface soils in Hong Kong.

Soil type	Area (%)	Landslide pixels (%)	Shear strength parameters	
			c (kPa)	f (°)
Completely decomposed granites	22.46	39.55	5–15	35–44
Completely decomposed volcanics (tuffs and rhyolites)	45.04	38.20	5–10	32–48
Colluvium (matrix material)	32.50	22.25	0–10	26–40

Note: the relatively wide ranges of the parameters reflect the variable composition of colluvial matrix material in Hong Kong.

Table 4
Description of land use.

Level	Category	Area (%)	Landslide pixel (%)	Description
1	Water body	1.14	0.02	Streams, nullahs, reservoirs and ponds
2	Forest	4.25	0.74	Woodland and shrubland
3	Grass	48.73	42.08	Grassland, mangrove and swamp
4	Agriculture	16.68	18.26	Agriculture land, fish pods
5	Industry	2.05	0.91	Industrial estates, science and technology parks
6	Transportation	5.84	2.09	Roads and transport facilities, railways and airport, port facilities
7	Residential	21.32	35.90	Private, public and rural residential, govern, institutional and community facilities

These mitigation strategies generally include strengthening the source area of landslides, such as upgrading hillside catchments, upgrading man-made slopes, and controlling the areas of historical landslides. The upgraded area for 2019 and the recent landslides from 2020 to 2021 (GEO, 2000) show that the recent landslides rarely occur in upgraded areas (Fig. 4c). In other words, the upgraded areas have a lower probability of landslide occurrence. Therefore, during the prediction of landslide probability, the effects of upgraded areas to landslide susceptibility should be considered. Previous studies have not paid attention to this prominent aspect of landslide susceptibility prediction (Ng et al., 2021a, 2021b; Xiao et al., 2022). Reasonable consideration of the influence of mitigation strategies is critical for predicting the regional susceptibility of landslides, thereby enhancing risk management capabilities.

3. Methodology

The proposed ML framework for landslide prediction considering mitigation strategies for landslide control includes three main parts: data acquisition and processing, ML modelling, and landslide susceptibility prediction (Fig. 6). The first part is data acquisition and processing, which involves collecting landslide influencing factors and data, which includes upgraded areas with historical landslides. In Section 2, landslide influencing factors were classified into rainfall, topographical, geological, and anthropogenic conditions. In this study, a total of 13 influencing factors are used for landslide susceptibility prediction—namely AAR, elevation, slope, plane curvature, profile curvature, aspect, TWI, road network density, river system density, geology, land use, soil properties (cohesion and friction angle), and influences of infrastructures. In addition, records of historical landslides and upgraded landslide areas form the basis for establishing ML models. The landslide pixels were used to extract information on the influencing factors, whereas non-landslide pixels were randomly produced with information on the influencing factors. It should be noted that the upgraded areas were used to generate additional non-landslide pixels to reflect the effects of mitigation strategies on landslide susceptibility. To consider the effects of mitigation strategies for landslide control on landslide susceptibility, two scenarios are considered. In Scenario I, landslide control is not considered whereas in Scenario II, landslide control is considered. Both scenarios were simulated and compared, and landslide and non-landslide pixels were used to produce a dataset that was split into training and test subsets in a 7:3 ratio.

Based on the established datasets, the next step involves training ML models in order to predict landslide susceptibility. In this study, the RF, LR, and SVC ML models were adopted. The performances of the different models were evaluated and compared using performance metrics, including the accuracy, recall, precision, F1-score, the receiver operator characteristic (ROC) curve, and the area under curve (AUC). Landslides from 1985 to 2021 were used to train the ML models. The trained models were utilized to predict the spatial distribution of landslide probabilities based on the GIS platform. Finally, historical landslides were used to validate the predicted results.

3.1. Determination of influencing factors

The influencing factors are important for the performance of ML models because they affect the reliability of the predictions. The determination of robust influencing factors is critical to obtain an acceptable prediction result. In order to support the selected influencing factors, previous studies on landslide susceptibility assessments were investigated. Table 5 lists the explanatory features used in previous studies. To avoid a collinearity problem with multiple determined influencing factors, pairwise correlations among the determined influencing factors on landslide susceptibility were conducted in Fig. 7. The values in the pairwise correlation matrix can reflect the relationships among these factors. The values close to 1 refer to a strong positive correlation between two influencing factors, while the values close to 0 indicate an independent relationship between two influencing factors. Based on the previous studies, the values excess than 0.7 may lead to a multicollinearity issue in the dataset (Chen et al., 2018). As shown in Fig. 7, the correlation values of all the influencing factors are less than 0.7. The result indicates that the determined influencing factors are independent each other.

3.2. ML models

In this study, three ML models, namely RF, LR, and SVC were used to predict landslide susceptibility. The prediction of landslide susceptibility is considered as a binary classification problem, with each sample resulting in either a positive (landslide) or negative (non-landslide) prediction. The analysis was conducted by applying the Python programming environment and ML package scikit-learn.

3.2.1. Random forest

RF is a popular ML approach for classification and regression problems that operates on more than one decision tree. The input variables are classified into a large number of random trees, and the output is the class determined by the largest selection. During training, RF produces a multitude of trees, and each tree is trained according to a random subset of input data. When conducting predictions, each tree yields its own decision, and the final decision is determined by the unweighted majority vote from all trees. RF is one of the preferred methods because of its high accuracy and speed (Ng et al., 2021a, 2021b; Wang et al., 2022).

3.2.2. Logistic regression

LR is one of the most widely used statistical ML methods for classification problems. The basic idea behind LR is to produce discrete binary outputs using a linear regression formula that expresses a dependent variable in terms of several independent variables (Xiao et al., 2022). Before the application of LR, the first step is the standardisation of the input variables, which can

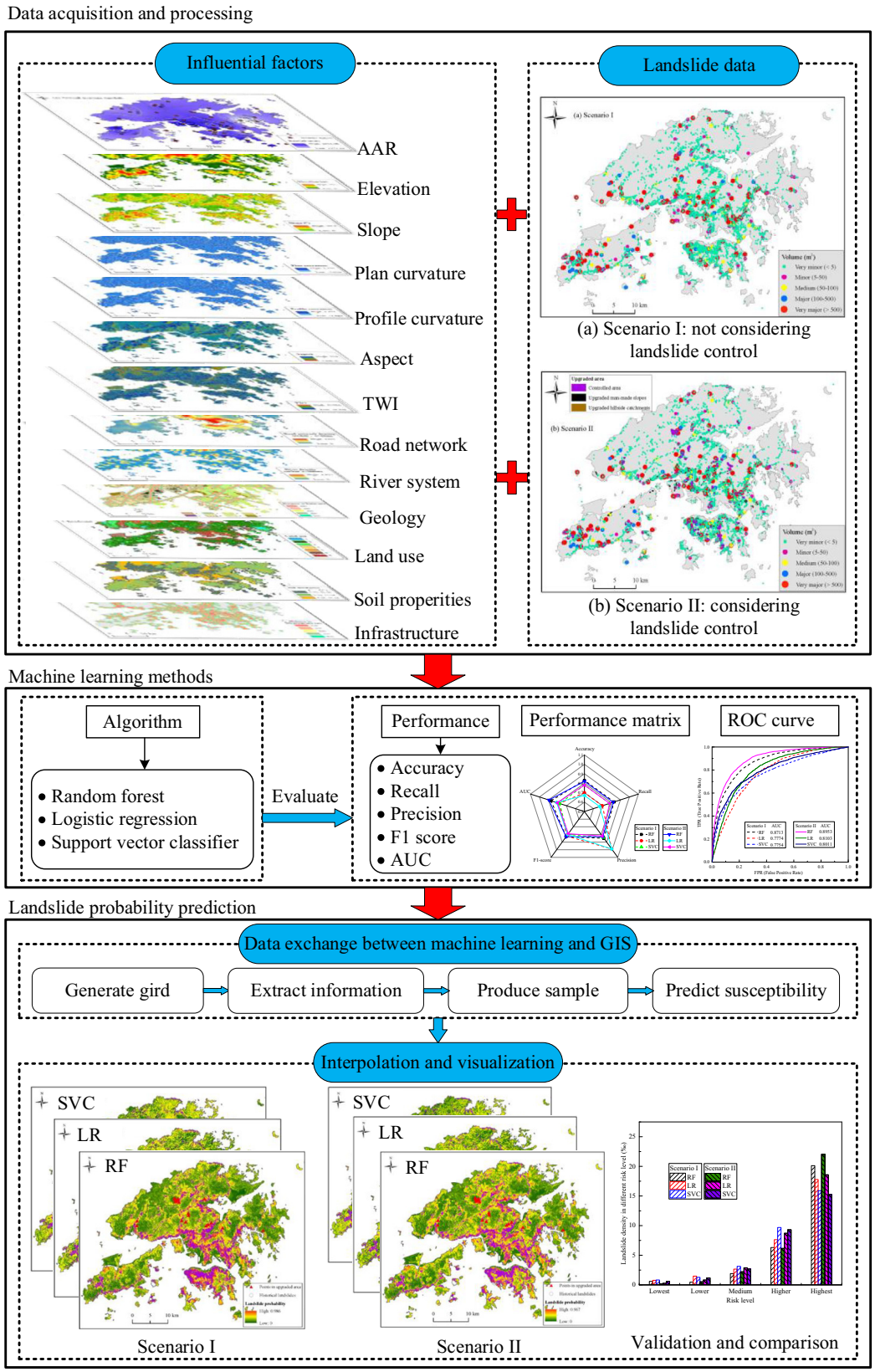

Fig. 6. Machine learning (ML) framework for landside prediction.

decrease the numerical differences among the input features, resulting in better predictions. In this study, the information of the features x_{itrain} from the training samples was used to standardise the features x_{itext} from the test sample, as follows:

$$x_{itest} = \frac{x_{itrain} - \bar{x}}{\sqrt{\frac{\sum_{i=1}^{n}\left(x_{itrain} - \bar{x}\right)^2}{n-1}}} \quad (2)$$

Table 5
Assessment factors adopted in various studies on landslide susceptibility.

	Annual average rainfall	Elevation	Slope	Plane curvature	Profile curvature	Aspect	TWI	Road network density	River system density	Geology	Land use type	Soil properties	Distance to Infrastructures
Huang et al., 2023		√	√	√	√	√	√	√	√		√	√	√
Nwazelibe et al., 2023		√	√	√	√	√	√	√	√	√	√		√
Tyagi et al., 2023	√		√				√	√	√		√	√	√
Xiao et al., 2022	√		√			√				√	√		√
Ullah et al., 2022	√	√	√	√	√	√	√	√	√	√	√		√
Li et al., 2022	√	√	√	√	√	√	√		√		√		
Ng et al., 2021a	√	√	√	√	√	√					√		
Salehpour Jam et al., 2021	√	√	√			√		√	√	√			√
Wang et al., 2021b	√	√	√	√	√	√	√	√	√			√	√
Wubalem, 2021			√	√	√	√			√	√	√	√	
Sun et al., 2020	√	√	√	√	√	√	√	√	√		√	√	√
Yi et al., 2020		√	√			√		√	√		√	√	√
Hong et al., 2018	√	√	√	√	√	√	√	√	√		√		√
Chen et al., 2018	√	√	√	√	√	√	√	√	√		√		√
Chen et al., 2017	√	√	√	√	√	√		√	√		√		√
Huang et al., 2017		√	√	√	√	√				√			

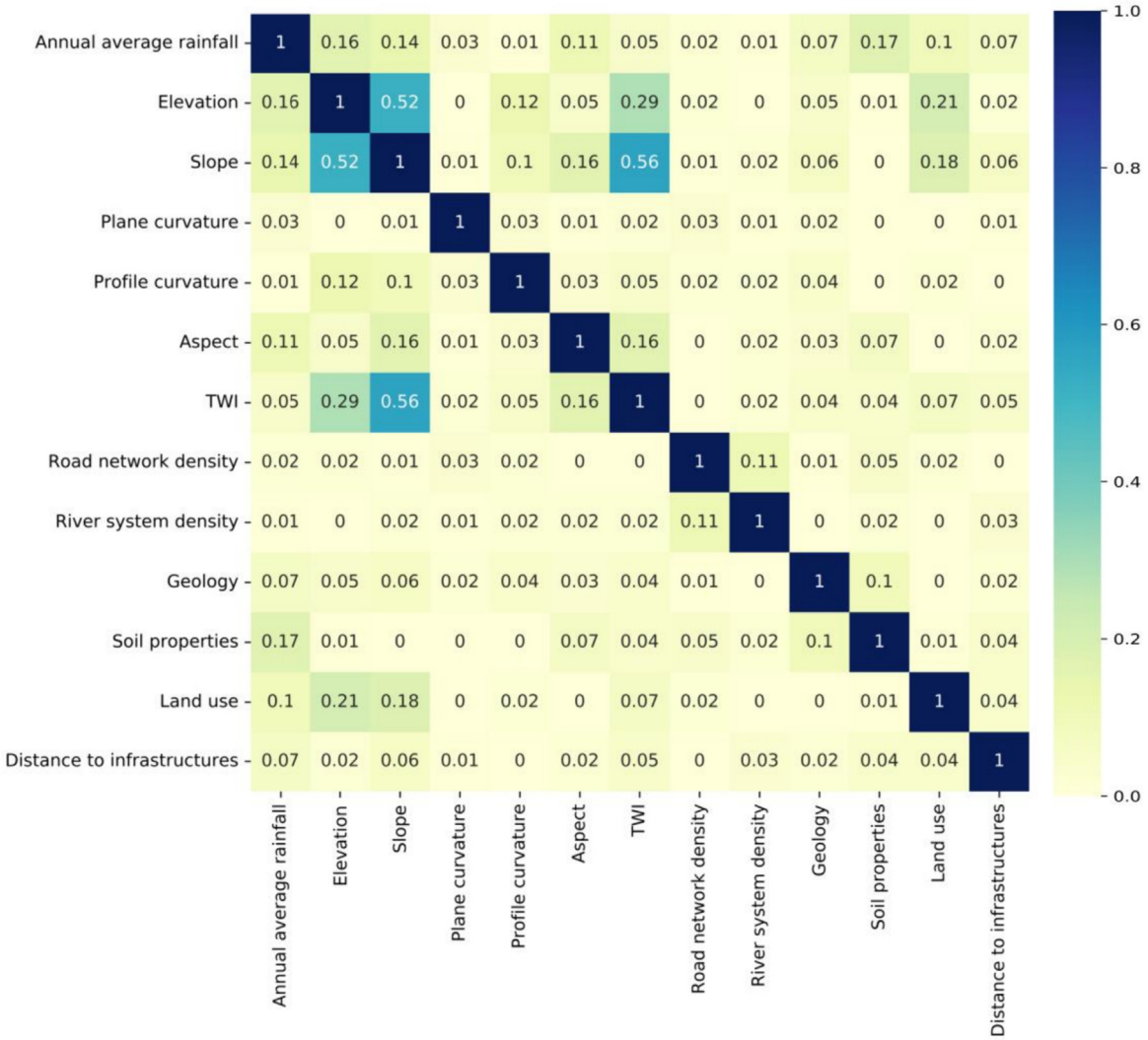

Fig. 7. Pairwise correlations among selected influencing factors.

where $\bar{x}$ is the mean value of the training sample, n is the number of test sample. This process allows the training samples to remember the information from the test samples. To achieve a better prediction, the input variables were processed with polynomial analysis. LR is suitable for landslide probability prediction since it has the ability to process both numerical and categorical variables (Budimir et al., 2015; Xiao et al., 2022).

3.2.3. Support vector classification

SVMs are another classical and popular ML method that can be divided into support vector regression (SVR) and support vector classification (SVC). The SVC is used to solve classification issues and is widely applied in various fields due to its ability to deal well with complex classification problems. SVC aims to find a linear hyperplane created by support vectors, which can be divided into negative and positive samples with the maximum margin. The kernel functions play a significant role in SVC by bridging linear and nonlinear problems using nonlinear decision boundaries. During the application of SVC, the same standardisation defined by Eq. (2) was used to standardise the input variables.

3.3. Difference between scenarios I and II

To demonstrate the efficiency of the proposed approach, both scenarios I and II were simulated and compared to consider the effects of mitigation strategies on landslide susceptibility. Scenario I that includes historical landslides from 1985 to 2021 (including 9843 landslide and 19,686 non-landslide samples) was envisaged as not considering mitigation strategies. The landslide samples are usually fewer, maybe half as many as the non-landslide samples in this study. The Scenario II was hypothesised as considering the mitigation strategies in the upgraded landslide area. To guarantee the accuracy of the trained models, the number of the samples in Scenario II is same as in Scenario I. To evaluate the effects of mitigation strategies on landslide susceptibility, non-landslide samples were produced in the upgraded area and added to the randomly created samples to serve as ML models in training datasets. The upgraded area includes the upgraded hillside catchments, upgraded man-made slopes and controlled area (Fig. 4c, Table 6). A partially enlarged schematic of the datasets and the classified study area is presented in Fig. 8, while Table 7 depicts the difference between scenarios I and II. Here, in Scenario I, the non-landslides samples are produced as double as landslide samples across the study area. In Scenario II, the non-landslide samples include two parts, which are 13,124 (two-thirds of the total number of landslides) non-landslides in study area and 6562 (one third of the total number of landslides) non-landslides in the upgraded area.

Table 6
Description of upgraded area with non-landslide samples.

Upgraded area	Area (km^2)	Non-landslide sample	
		Scenario I	Scenario II
Upgraded hillside catchments	4.146	142	854
Upgraded man-made slopes	9.267	318	1908
Controlled area	18.454	633	3800

3.4. Data exchange between ML and GIS

Both landslide and non-landslide samples were applied to extract information on the control indices, which produced GIS databases. The databases containing information on the factors influencing landslides and non-landslides were converted into CSV files for the implementation of the ML methods. Based on training and testing from ML modelling, the best models from these two scenarios were selected to predict the landslide probabilities of the grid points in the study area. The study area was divided into 118,008 grids with a size of 100 m × 100 m using a fishnet created in GIS. Generally, a 30 m resolution is used to generate landslide susceptibility, and the accuracy of landslide susceptibility increases with a relatively smaller grid size (Liao et al., 2022). However, with a 30 m resolution, the number of grids exceed 400,000 in the study area. Classifying the study area with 30 m grids to compute the landslide probabilities was also attempted, but it is difficult to process such a large amount of data. In the previous publication from Xiao et al. (2022), the grids were classified with a size of 750 m × 600 m to predict the spatial distribution of slopes in Hong Kong. Moreover, it should be noted that the classified grid is not used to represent landslides. The landslide probability of each point in a specific grid was predicted, based on the probabilities between two adjacent points obtained by the Kriging interpolation analysis. The grids with centre points were used to conduct interpolation analysis to obtain landslide susceptibility maps.

4. Model evaluation and analysis

4.1. Tuning of hyper-parameters

Hyper-parameters are referred to as the non-constant variables in ML models. To obtain the optimum prediction from the ML models, the hyper-parameters should be firstly tuned to generate the optimum model. The hyper-parameters for RF, LR, and SVC were determined using a grid search approach based on a 10-fold cross-validation of the training samples. The hyper-parameters of the classifier models for the two considered scenarios are listed

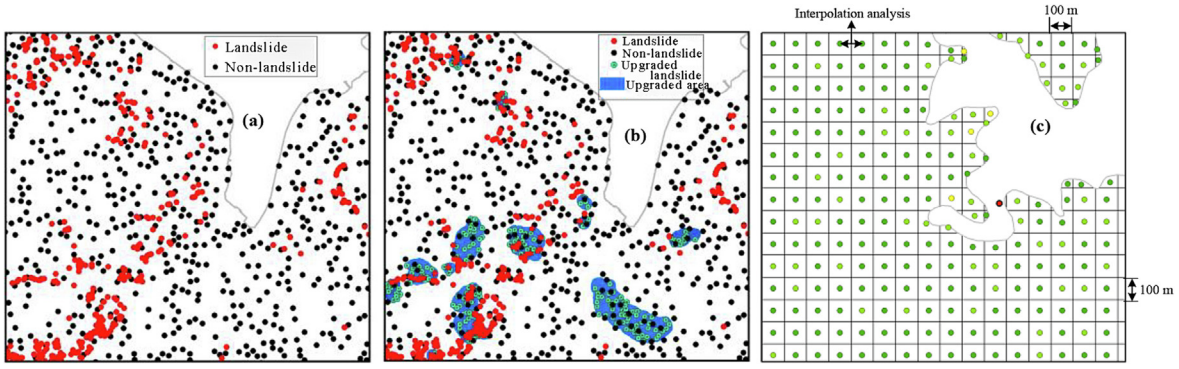

Fig. 8. Partially enlarged schematic: (a) Scenario I; (b) Scenario II (considering landslide control); (c) fishnet grid with points in the study area.

Table 7
Difference between Scenarios I and II.

Scenarios	Landslide samples	Non-landslide samples
I	9843	19,686 (double landslides) in study region without upgraded area
II	9843	13,124 (two-thirds of landslides) in study area and 6562 (one third of landslides) in upgraded area

in Table 8. The input variables for the LR and SVC were processed through standardisation, prior to training the models. The RF, LR, and SVC were retrained after determining the optimum hyperparameter combinations.

4.2. Performance evaluation

Performance metrics, such as accuracy, recall, precision, F1-score, ROC curve, and AUC, were used to evaluate and compare the performances of the different ML models. The number of landslides classified as landslides was defined as true positive (TP), the number of non-landslides classified as non-landslides corresponding to true negative (TN), the number of landslides classified as non-landslides was defined as false positive (FN), and the number of non-landslides classified as landslides was false negative (FP). A detailed definition of the evaluation of performance metrics can be found in previous publications (Cheung, 2021; Ng et al., 2021a, 2021b).

The ROC curve, which compares the true positive rate (TPR) with the false positive rate (FPR), is used to measure the performance of a binary classifier. The TPR is the ratio of TP to all actual positive samples (TP + FN), whereas the FPR is the ratio of FP to all actual negative samples (TN + FP). The AUC of the ROC curve provides an overall evaluation of model performance. The greater the AUC, the better the model performance. The performance evaluation of machine learning models includes the ROC curves (Fig. 9a) and performance matrixes (Fig. 9b). Based on the performance evaluation, the RF model appeared to be the optimum, with test AUC values of 0.8713 and 0.8953 for scenarios I and II, respectively, followed by the LR (0.7774 and 0.8103 for scenarios I and II, respectively), and SVC (0.7754 and 0.8011 for scenarios I and II, respectively). Compared to other landslide susceptibility studies, differences in performance are apparent. Some ML models achieved AUC values above 0.98 (Xiao et al., 2022). However, the performance metrics are in a similar range as in the study by Ng et al., (2021a, 2021b), whose winning model achieved an AUC value of 0.89, and an accuracy of 0.82. With the spatially homogeneous distribution of the sample data, the ML models achieved acceptable performance. Fig. 9b shows the accuracy, recall, precision, F1-score, and AUC. It can be seen that the accuracy of RF is the highest, followed by LR and SVC in scenarios I and II. RF yielded the highest recall, precision, and F1-score, followed by LR and SVC. Moreover, a comparison between scenarios I and II showed that the test performance metrics for Scenario II were better than those of Scenario I. The models trained in Scenario II performed better than those trained in Scenario I. This indicates that mitigation strategies should be considered for landslide control.

4.3. Predicted landslide susceptibility

The RF, LR, and SVC models were implemented with 13 landslide influencing factors to produce landslide susceptibility maps for scenarios I and II in Hong Kong. Fig. 10 depicts the spatial distribution of landslide susceptibilities in Scenario I as predicted by the RF (Fig. 10a), LR (Fig. 10b), and SVC (Fig. 10c) ML models. As shown in Fig. 10, the landslide susceptibility predicted by the RF ranges from 0 to 0.986, followed by the LR (from 0 to 0.955) and SVC (from 0 to 0.947). Fig. 11 shows the spatial distribution of landslide susceptibilities in Scenario II as predicted by the RF (Fig. 11a), LR (Fig. 11b), and SVC (Fig. 11c). As shown in Fig. 11, the landslide susceptibility predicted by the RF ranged from 0 to 0.967, followed by the LR (from 0 to 0.936) and SVC (from 0 to 0.902). The comparison between these two scenarios indicates that the largest susceptibilities predicted in Scenario II are less than that in Scenario I. This implies that when considering mitigation strategies for landslide control in the prediction, the landslide susceptibility magnitudes tended to decrease. This phenomenon is further illustrated at the locations of historical landslides (Figs. 10 and 11), where most historical landslides were located within the areas identified by the RF model as high landslide susceptibility zone. Areas with high landslide susceptibilities were located on Hong Kong Island, the Kowloon Peninsula, and the south of the New Territories, where the buildings and population are densely concentrated. Moreover, the largest landslide susceptibility values were predicted as 0.967 in RF, 0.936 in LR and 0.902 in SVC at point B (shown in Fig. 11), where the Tung Lo Wan landslide in 2008 were happened. These maps indicate the risk of landslides at specific locations depending on the corresponding influencing factors. Therefore, it is necessary to analyse the contribution of each influencing factor to landslide susceptibility.

4.4. Contribution from influencing factors

The contribution of the factors influencing landslide susceptibility show that the AAR has the largest contribution (11.06%, with a fluctuation range of 1.2%) (Fig. 12). Rainfall provided greater contributions than other control indices. Topographical conditions such as elevation, slope, aspect, TWI, road network and river system densities also contributed significantly to landslide susceptibility. Soil mechanical properties had a relatively smaller contribution to landslide susceptibility than the other controlling

Table 8
Tuned hyperparameters of each classifier machine learning model.

Algorithm	Hyperparameters	Definition	Value	
			Scenario I	Scenario II
RF	*n_estimators*	The number of decision trees in the forest	125	150
	min_samples_leaf	The number of minimum samples leaf in the forest	10	10
	max_depth	The maximum depth for a decision tree can obtain	19	21
	criterion	The function for determining the quality of a split	entropy	entropy
	class_weight	The way to reach a best classifier	balanced	balanced
	learning_rate	The contribution of each decision tree in the forest	1	1
LR	C_s	The penalty parameter determined the tolerance of error	0.5	10
	penalty	The norm of penalty (Lasso regression)	L2	L2
	degree	Features of polynomial	5	4
SVC	*gamma*	Influences of a plausible line of calculation	0.001	0.0001

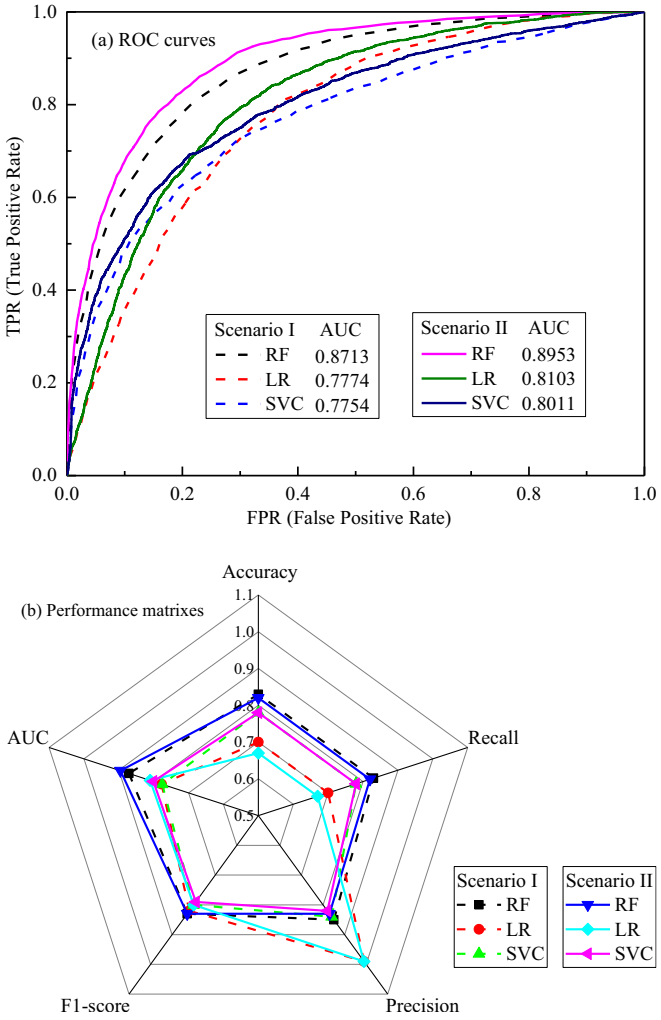

Fig. 9. Performance evaluation of machine learning models: (a) ROC curves; (b) performance matrixes.

factors influencing infrastructure exhibited the largest fluctuation range (1.7%) because the influence of infrastructure was quantified by distances of 100, 200, 300, 400, and 500 m to buildings and metro lines. This type of infrastructure value results in a fluctuating contribution to landslide susceptibility.

5. Discussion

5.1. Validation

5.1.1. Validation against historical landslides

The generated landslide susceptibility was classified into five risk levels, from very low to very high, to reflect the risk to trigger landslides. Previous studies have used several methods to classify landslide susceptibility, such as quantiles, natural breaks, equal intervals, and standard deviations. The natural break classification method is widely used. The ratios of the area corresponding to different landslide susceptibility levels according to the results obtained from the RF, LR, and SVC models in scenarios I and II are shown in Fig. 13. Here, the area classified as having the highest landslide susceptibility level was the largest in the case of RF, followed by those from the LR and SVC models. Moreover, the ratios of the lowest level predicted from Scenario I were less than those from Scenario II, and the ratios of the highest level predicted from Scenario I were larger than those from Scenario II. The results indicate that Scenario II decreases the predicted landslide susceptibility by realistically considering the mitigation strategies, unlike Scenario I where the landslide susceptibility is overestimated.

The densities of historical landslides at different risk levels were determined to validate the predicted landslide susceptibility. Density, here, is defined as the ratio between the number of pixels at a specific risk level and the total number of pixels at a specific risk level, as illustrated in Eq. (3). The density of historical landslides in areas with different risk levels, where the highest risk typically has the largest density of historical landslides, has the highest value in the case of the RF, followed by LR and SVC in both, scenarios I and II (Fig. 14). This implies that most historical landslides were located in areas with the highest risk. Generally, if more landslides are reported in high-risk areas and fewer landslides are reported from low-risk areas, the better the prediction ability of the model. Therefore, the results indicate an acceptable prediction ability of the proposed model based on machine learning and GIS.

$$\text{Density} = \frac{\text{Number of pixels at specific risk level}}{\text{Total pixels at specific risk level}} \quad (3)$$

5.1.2. Validation against points in upgraded areas

The landslide probabilities of the points in the upgraded areas were extracted to reflect the effects of the mitigation strategies on landslide control. The landslide probabilities of 20 points randomly produced in the upgraded area in scenarios I and II are shown in Fig. 15, where for the same point, the landslide probabilities predicted from Scenario II are less than those from Scenario I when using the same ML model. The comparison between scenarios I and II indicate that the consideration of mitigation strategies for landslide control can realistically decrease landslide susceptibility. As part of the management of landslide susceptibility in the region, various countermeasures have been adopted to control landslide hazards, which substantiate the observed decreasing trend in predicted landslide numbers. The results from Scenario II demonstrate the pertinence of predicting landslide probability by considering mitigation strategies.

5.2. Guidance for landslide management

Areas with high landslide susceptibility tend to be located on the Hong Kong Island, the Kowloon Peninsula, and the south of the New Territories, which have a dense distribution of both, buildings and population. Strict and scientific landslide control measures, such as landslide monitoring and detection, should be implemented in areas with high landslide susceptibility (Cheung, 2021; Ng et al., 2021a, 2021b). The AAR is the most important driving factor contributing to landslide susceptibility. With a potential increase in rainfall intensity due to climate change in the coming years, it is essential to improve rainfall forecasting and early warning for the spatial distribution of rainfall. In this respect, the GEO operates on an extensive network of automatic rain gauges that provide real-time rainfall data to support landslide warning systems. Moreover, factors that are influenced by topography are also responsible in contributing to landslide susceptibility; thus, a detailed natural terrain inventory should be implemented to identify and assess sites that are prone to landslides (Ng et al., 2003).

5.3. Limitations

The prediction method based on ML and GIS in this study can be easily implemented in other landslide-prone areas to develop regional and spatial landslide prediction models. A major obstacle is data scarcity. Compared with the relatively abundant and easy-to-obtain data on rainfall and geological conditions, databases for

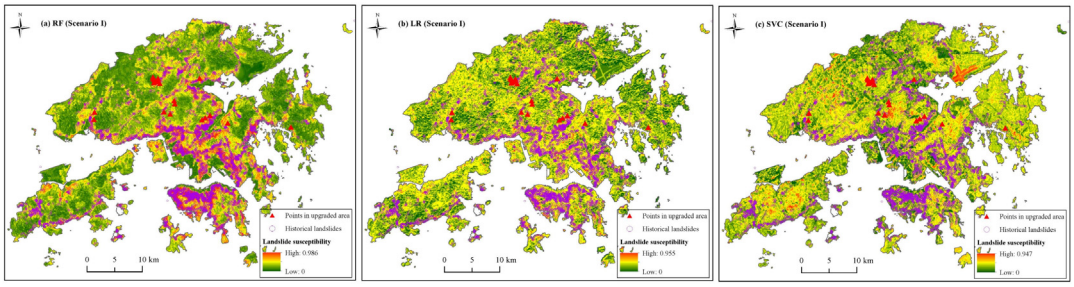

Fig. 10. Landslide susceptibility maps in Scenario I predicted by (a) RF, (b) LR, and (c) SVC.

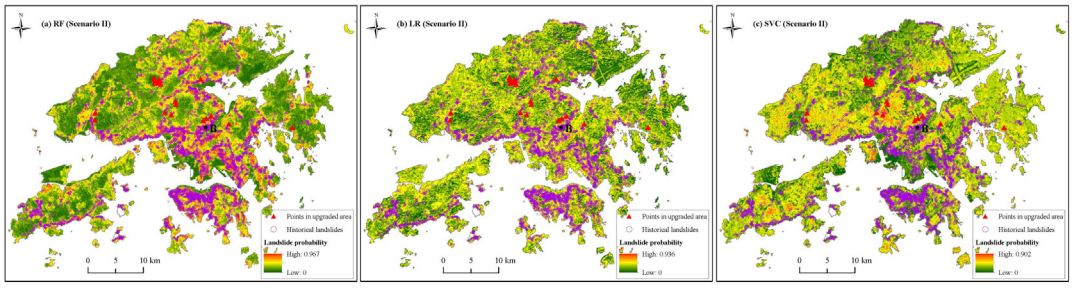

Fig. 11. Landslide susceptibility maps in Scenario II predicted by (a) RF, (b) LR, and (c) SVC.

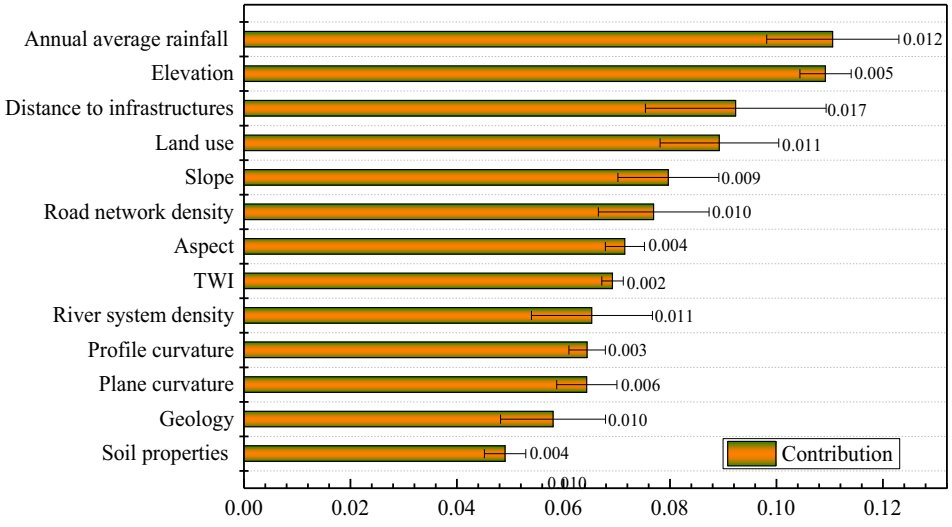

Fig. 12. Contribution of the influencing factors to landslide susceptibility.

landslides or the categories of engineering slopes are limited and not always available, which may limit the application of the proposed approach. Furthermore, assessing landslide susceptibility is complex due to the combined effects of natural and social factors. Thus, the selected 13 influencing factors may be insufficient to fully reflect the factors controlling landslide susceptibility. For instance, rather than remaining passive, anthropogenic activities can resist and adapt to natural hazards. However, this study did not consider influencing factors that qualify as countermeasures or social and economic resilience, because they cannot be clearly defined; for instance, mitigation strategies have not been considered an influencing factor in this study. The mitigation strategies are reflected by the difference of non-landslide samples in scenarios I and II. The number of non-landslide samples may lead to uncertainties of results. Moreover, the study area was classified with 100 m grids, and each grid with a center point (Fig. 16). The landslide probability of each point was predicted, after which the probabilities between two adjacent points were obtained by interpolation analysis. By using the interpolation analysis, we can obtain the spatial distribution of landslide susceptibilities. This process may induce uncertainties in the results. In addition, the hyper-parameter values of all ML models cannot cover all possible conditions because they are constrained to empirical ranges, resulting in an unfair model comparison.

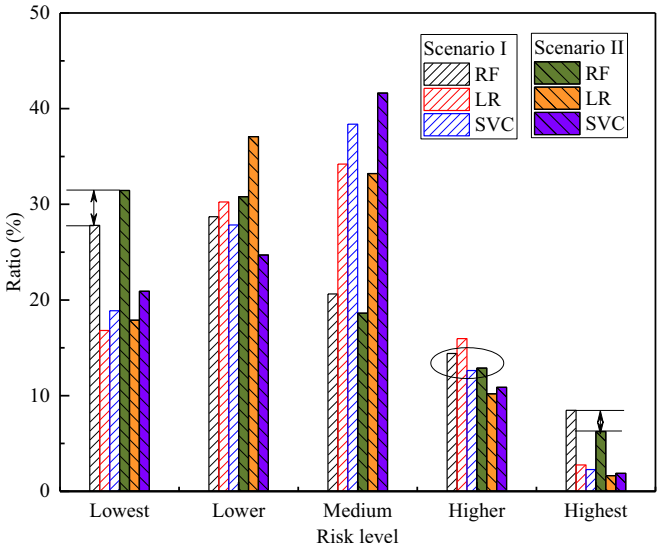

Fig. 13. Percentages of area corresponding to different landslide risk levels.

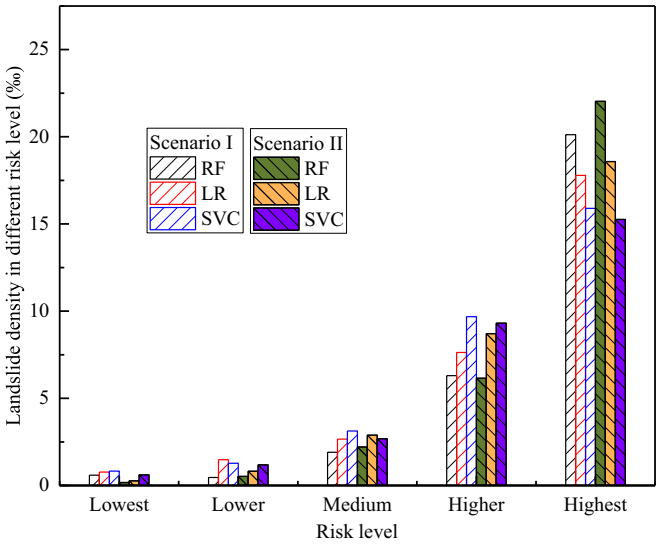

Fig. 14. Density of historical landslides in different risk levels.

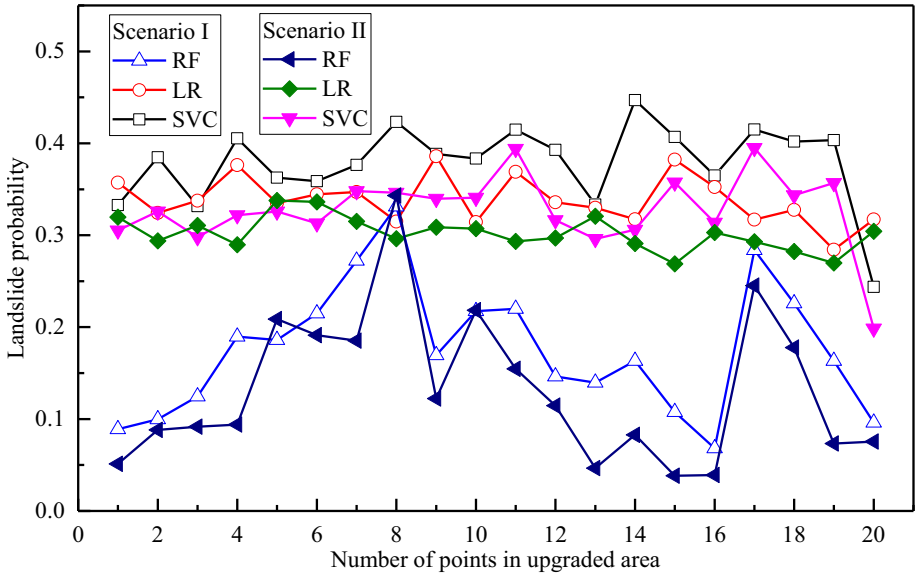

Fig. 15. Landslide probabilities of points in upgraded area in scenarios I and II.

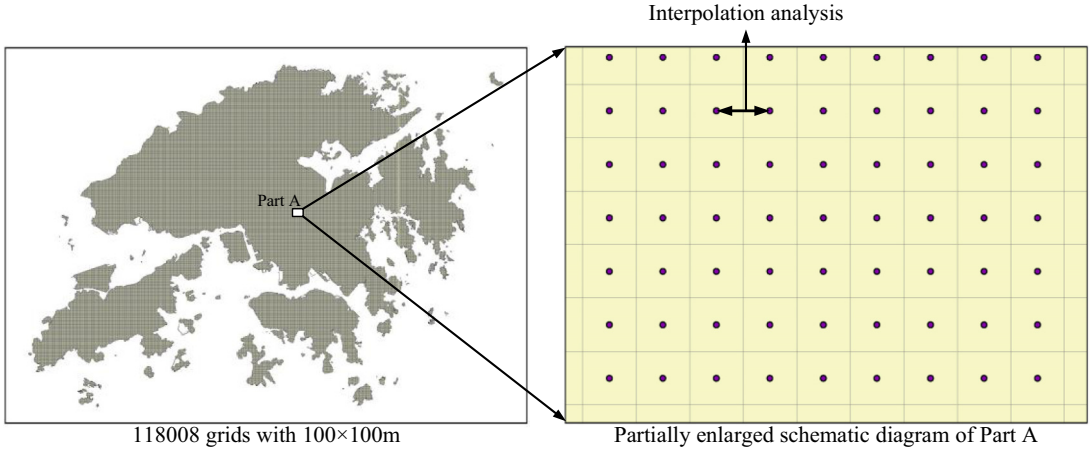

Fig. 16. Schematic of the classified grids for study area.

6. Conclusions

This study proposes an approach to consider mitigation strategies on landslide susceptibility prediction in ML models. An integrated approach, which incorporates ML models (e.g. RF, LR, and SVC) into GIS was adopted to predict the spatial distribution of landslide probabilities in Hong Kong. Landslide influencing factors including rainfall, topography, geology, anthropogenic conditions, and mechanical properties of soil, were used to establish a database within a GIS platform. To consider the effects of the mitigation strategies for landslide control on landslide susceptibility, two scenarios were simulated and compared; Scenario I does not consider control and Scenario II considers landslide control. The major conclusions drawn from this study are summarised as follows.

(1) The landslide susceptibilities were predicted by considering landslide mitigation strategies. The mitigation strategies were reflected by the difference of non-landside samples in scenarios I and II. The effects of mitigation strategies for landslide control were evaluated by producing non-landslide samples in the upgraded area as well as randomly created samples to serve as ML model training datasets.
(2) The comparison between scenarios I and II indicate that mitigation strategies for landslide control could decrease landslide susceptibility. When considering landslide control, the landslide susceptibilities predicted in Scenario II were lower than those obtained from Scenario I. This invoked interest in considering mitigation strategies within the ML models for landslide control.
(3) Among the considered landslide controlling factors, rainfall contributed the most to landslide probability prediction, whereas soil mechanical properties contributed the least. This is mainly because shallow soil is vulnerable to rainstorms as its strength decreases with increasing water content.
(4) The RF, LR, and SVC models were incorporated into the GIS to predict landslide susceptibilities in scenarios I and II in Hong Kong. Validation against historical landslides demonstrated that RF performed the best, followed by LR and SVC, in predicting landslide susceptibilities. The integration of ML and GIS provide excellent predictions of the spatial distribution of landslide susceptibility, making it a powerful tool for landslide susceptibility management.

CRediT authorship contribution statement

Hai-Min Lyu: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. **Zhen-Yu Yin:** Conceptualization, Supervision, Validation, Writing – review & editing. **Pierre-Yves Hicher:** Formal analysis, Validation, Visualization, Writing – review & editing. **Farid Laouafa:** Formal analysis, Validation, Visualization, Writing – review & editing.

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

The authors gratefully acknowledge funding by the National Natural Science Foundation of China (Grant No. 42007416), the Hong Kong Polytechnic University Strategic Importance Fund (ZE2T) and Project of Research Institute of Land and Space (CD78).

References

Abraham, M.T., Vaddapally, M., Satyam, N., Pradhan, B., 2023. Spatio-temporal landslide forecasting using process-based and data-driven approaches: A case study from Western Ghats, India. Catena 223, 106948. https://doi.org/10.1016/j.catena.2023.106948.
Al-Najjar, H.A.H., Pradhan, B., Beydoun, G., Sarkar, R., Park, H.-J., Alamri, A., 2023. A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset. Gondwana Res. 123, 107–124. https://doi.org/10.1016/j.gr.2022.08.004.
Budimir, M.E.A., Atkinson, P.M., Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides 12, 419–436. https://doi.org/10.1007/s10346-014-0550-5.
Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Zhou, D., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151, 147–160. https://doi.org/10.1016/j.catena.2016.11.032.
Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A.X., Pei, X., Duan, Z., 2018. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Sci. Total Environ. 626, 1121–1135. https://doi.org/10.1016/j.scitotenv.2018.01.124.
Cheung, R.W.M., 2021. Landslide risk management in Hong Kong. Landslides 18, 3457–3473. https://doi.org/10.1007/s10346-020-01587-0.
Dematteis, N., Wrzesniak, A., Allasia, P., Bertolo, D., Giordan, D., 2022. Integration of robotic total station and digital image correlation to assess the three-dimensional surface kinematics of a landslide. Eng. Geol. 303, 106655. https://doi.org/10.1016/j.enggeo.2022.106655.
Feng, W., Bai, H., Lan, B., Wu, Y., Wu, Z., Yan, L., Ma, X., 2022. Spatial–temporal distribution and failure mechanism of group-occurring landslides in Mibei village, Longchuan County, Guangdong, China. Landslides 19, 1957–1970. https://doi.org/10.1007/s10346-022-01904-9.
GEO, 2000. Guide to Retaining Wall Design. Geotechnical Engineering Office (GEO), Civil Engineering Department, The Government of the Hong Kong Special Administrative Region, 137-140 pp.
Hong, H., Liu, J., Bui, D.T., Pradhan, B., Acharya, T.D., Pham, B.T., Zhu, A., Chen, W., Ahmad, B.B., 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena 163, 399–413. https://doi.org/10.1016/j.catena.2018.01.005.
Huang, W., Ding, M., Li, Z., Yu, J., Ge, D., Liu, Q., Yang, J., 2023. Landslide susceptibility mapping and dynamic response along the Sichuan-Tibet transportation corridor using deep learning algorithms. Catena 222, 106866. https://doi.org/10.1016/j.catena.2022.106866.
Huang, F., Yin, K., Huang, J., Gui, L., Wang, P., 2017. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. Eng. Geol. 223, 11–22. https://doi.org/10.1016/j.enggeo.2017.04.013.
Ju, L.Y., Xiao, T., He, J., Wang, H.J., Zhang, L.M., 2022. Predicting landslide runout paths using terrain matching-targeted machine learning. Eng. Geol. 311, 106902. https://doi.org/10.1016/j.enggeo.2022.106902.
Khezri, S., Ahmadi Dehrashid, A., Nasrollahizadeh, B., Moayedi, H., Dehrashid, H.A., Azadi, H., Scheffran, J., 2022. Prediction of landslides by machine learning algorithms and statistical methods in Iran. Environ. Earth Sci. 81, 304. https://doi.org/10.1007/s12665-022-10388-8.
Lam, C.L.H., Lau, J.W.C., Chan, H.W., 2012. Factual report on Hong Kong rainfall and landslides in 2008. GEO Report 273, Geotechnical Engineering Office, Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region, 87-95pp.
Li, W., Fang, Z., Wang, Y., 2022. Stacking ensemble of deep learning methods for landslide susceptibility mapping in the Three Gorges Reservoir area, China. Stoch. Environ. Res. Risk Assess. 36, 2207–2228. https://doi.org/10.1007/s00477-021-02032-x.
Liao, M., Wen, H., Yang, L., 2022. Identifying the essential conditioning factors of landslide susceptibility models under different grid resolutions using hybrid machine learning: A case of Wushan and Wuxi counties, China. Catena 217, 106428. https://doi.org/10.1016/j.catena.2022.106428.
Liu, Z., Gilbert, G., Cepeda, J.M., Lysdahl, A.O.K., Piciullo, L., Hefre, H., Lacasse, S., 2020. Modelling of shallow landslides with machine learning algorithms. Geosci. Front. 12 (1), 385–393. https://doi.org/10.1016/j.gsf.2020.04.014.
Lombardo, L., Tanyas, H., Huser, R., Guzzetti, F., Castro-Camilo, D., 2021. Landslide size matters: A new data-driven, spatial prototype. Eng. Geol. 293, 106288. https://doi.org/10.1016/j.enggeo.2021.106288.
Lyu, H.M., Yin, Z.Y., 2023. An improved MCDM combined with GIS for risk assessment of multi-hazards in Hong Kong. Sustain. Cities. Soc. 91, 104427. https://doi.org/10.1016/j.scs.2023.104427.
Lyu, H.M., Yin, Z.Y., Zhou, A., Shen, S.L., 2024. Sensitivity analysis of typhoon-induced floods in coastal cities using improved ANP-GIS. Int. J. Disast. Risk. Re. 104 (2024), 104344. https://doi.org/10.1016/j.ijdrr.2024.104344.
Ma, J., Xia, D., Wang, Y., Niu, X., Jiang, S., Liu, Z., Guo, H., 2022. A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: Insights from a case study of landslide displacement prediction. Eng. Appl. Artif. Intell. 114, 105150. https://doi.org/10.1016/j.engappai.2022.105150.
Ng, K.C., Parry, S., King, J.P., Franks, C.A.M., Shaw, R., 2003. Guidelines for Natural Terrain Hazard Studies, GEO Report No. 138., Geotechnical Engineering Office,

Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region, 176pp.
Ng, C.W.W., Yang, B., Liu, Z.Q., Kwan, J.S.H., Chen, L., 2021a. Spatiotemporal modelling of rainfall-induced landslides using machine learning. Landslides 18, 2499–2514. https://doi.org/10.1007/s10346-021-01662-0.
Ng, C.W.W., Liu, H., Choi, C.E., Kwan, J.S.H., Pun, W.K., 2021b. Impact dynamics of boulder-enriched debris flow on a rigid barrier. J. Geotech. Geoenviron. Eng. 147 (3), 04021004. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002485.
Nwazelibe, V.E., Unigwe, C.O., Egbueri, J.C., 2023. Testing the performances of different fuzzy overlay methods in GIS-based landslide susceptibility mapping of Udi Province, SE Nigeria. Catena 220 (A), 106654. https://doi.org/10.1016/j.catena.2022.106654.
Park, H.J., Lee, J.H., 2022. A review of quantitative landslide susceptibility analysis methods using physically based modelling. J. Eng. Geol. 32 (1), 27–40. https://doi.org/10.9720/kseg.2022.1.027.
Park, S.Y., Moon, S.W., Choi, J., Seo, Y.S., 2021. Machine-learning evaluation of factors influencing landslides. J. Eng. Geol. 31 (4), 701–718. https://doi.org/10.9720/kseg.2021.4.701.
Pradhan, B., Sameen, M.I., Al-Najjar, H.A.H., Sheng, D., Alamri, A.M., Park, H.-J., 2021. A meta-learning approach of optimisation for spatial prediction of landslides. Remote Sens. 13 (22), 4521. https://doi.org/10.3390/rs13224521.
Pradhan, B., Lee, S., Dikshit, A., Kim, H., 2023a. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. Geosci. Front. 14 (6), 101625. https://doi.org/10.1016/j.gsf.2023.101625.
Pradhan, B., Dikshit, A., Lee, S., Kim, H., 2023b. An explainable AI (XAI) model for landslide susceptibility modeling. Appl. Soft Comput. 142, 110324. https://doi.org/10.1016/j.asoc.2023.110324.
Salehpour Jam, A., Mosaffaie, J., Sarfaraz, F., Shadfar, S., Akhtari, R., 2021. GIS-based landslide susceptibility mapping using hybrid MCDM models. Nat. Hazards 108, 1025–1046. https://doi.org/10.1007/s11069-021-04718-5.
Shen, P., Wei, S., Shi, H., Gao, L., Zhou, W.H., 2023. Coastal flood risk and smart resilience evaluation under a changing climate. Ocean-Land-Atmos Res. 2, 0029 https://spj.science.org/doi/10.34133/olar.0029.
Sun, D., Wen, H., Wang, D., Xu, J., 2020. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362, 107201. https://doi.org/10.1016/j.geomorph.2020.107201.
Sun, D., Wen, H., Xu, J., Zhang, Y., Wang, D., Zhang, J., 2021. Improving geospatial agreement by hybrid optimization in logistic regression-based landslide susceptibility modelling. Front. Earth Sci. 9, 713803. https://doi.org/10.3389/feart.2021.713803.
Tyagi, A., Tiwari, R.K., James, N., 2021. GIS-based landslide hazard zonation and risk studies using MCDM. In: Sitharam, T.G., Jakka, R., Govindaraju, L. (Eds.), Local Site Effects and Ground Failures. Lecture Notes in Civil Engineering 117, Springer, Singapore, 251–266. https://doi.org/10.1007/978-981-15-9984-2_22.
Tyagi, A., Tiwari, R.K., James, N., 2023. Mapping the landslide susceptibility considering future land-use land-cover scenario. Landslides 20 (1), 65–76. https://doi.org/10.1007/s10346-022-01968-7.
Ullah, K., Wang, Y., Fang, Z., Wang, L., Rahman, M., 2022. Multi-hazard susceptibility mapping based on Convolutional Neural Networks. Geosci. Front. 13 (5), 101425. https://doi.org/10.1016/j.gsf.2022.101425.
Wang, X., Clague, J.J., Crosta, G.B., Sun, J., Stead, D., Qi, S., Zhang, L., 2021a. Relationship between the spatial distribution of landslides and rock mass strength, and implications for the driving mechanism of landslides in tectonically active mountain ranges. Eng. Geol. 292, 106281. https://doi.org/10.1016/j.enggeo.2021.106281.
Wang, Y., Tang, H., Huang, J., Wen, T., Ma, J., Zhang, J., 2022. A comparative study of different machine learning methods for reservoir landslide displacement prediction. Eng. Geol. 298, 106544. https://doi.org/10.1016/j.enggeo.2022.106544.
Wang, H., Zhang, L.M., Yin, K., Luo, H., Li, J., 2020. Landslide identification using machine learning. Geosci. Front. 12, 351–364. https://doi.org/10.1016/j.gsf.2020.02.012.
Wang, S., Zhuang, J., Zheng, J., Fan, H., Kong, J., Zhan, J., 2021b. Application of bayesian hyperparameter optimized random forest and xgboost model for landslide susceptibility mapping. Front. Earth Sci. 9, 712240. https://doi.org/10.3389/feart.2021.712240.
Wubalem, A., 2021. Landslide susceptibility mapping using statistical methods in Uatzau catchment area, northwestern Ethiopia. Geoenviron. Disasters 8, 1. https://doi.org/10.1186/s40677-020-00170-y.
Xiao, T., Zhang, L.M., Cheung, R.W.M., Lacasse, S., 2022. Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques. Geotechnique 73 (9), 749–765. https://doi.org/10.1680/jgeot.21.00160.
Yi, Y., Zhang, Z., Zhang, W., Jia, H., Zhang, J., 2020. Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. Catena 195, 104851. https://doi.org/10.1016/j.catena.2020.104851.