

Landslides (2024) 21:257–273  
 DOI 10.1007/s10346-023-02152-1  
 Received: 23 November 2022  
 Accepted: 16 September 2023  
 Published online: 17 October 2023  
 © Springer-Verlag GmbH Germany,  
 part of Springer Nature 2023

Tianhe Ren · Liang Gao  · Wenping Gong

# An ensemble of dynamic rainfall index and machine learning method for spatiotemporal landslide susceptibility modeling



**Abstract** Rain-induced landslides are one of the most recurrent geohazards around the world, posing great threats to the safety and property of people living in mountainous areas. One of the most effective strategies for rain-induced landslide risk management and reduction is to analyze landslide susceptibility during rain events. To characterize the impact of varying rainfall conditions on landslide occurrence, a maximum rolling rainfall index (MRRI) is proposed in this study for the spatiotemporal landslide susceptibility modeling. During rainfall, MRRI can be updated based on the real-time maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data. Furthermore, a spatiotemporal landslide susceptibility modeling approach, in which the MRRI index is taken as a conditioning factor, is developed. To illustrate the effectiveness and versatility of the proposed approach, landslide susceptibility models based on the random forest technique are trained in the central area of Hong Kong. The resulting landslide susceptibility models are then applied to two historical rainstorms, and the application results show that derived time-series landslide susceptibility maps are in good agreement with the spatial distribution of real landslides.

**Keywords** Landslide susceptibility · Rain-induced landslide · Spatiotemporal prediction · Maximum rolling rainfall index

## Introduction

Rain-induced landslides pose great threats to the safety and property of people living in mountainous areas around the world (Gong et al. 2021; Cheng et al. 2021). As the intensity and frequency of severe rainfall events tend to increase in most regions, landslide events will be more severe (Haque et al. 2019). Landslide risk should be controlled at a practicable level to meet the needs of public safety and sustainable development. Spatiotemporal prediction of rainfall-induced landslides plays a vital role in landslide risk management and mitigation in mountainous areas (Huang and Zhao 2018). Landslide susceptibility analysis, which can derive the spatial distribution of landslides given conditioning factors, is an effective approach for the spatial forecasting of landslides. Data-driven methods, particularly machine learning techniques, have been excessively adopted to build landslide susceptibility models. For example, logistic regression (Dai and Lee 2001; Sun et al. 2021), support vector machine (Yao et al. 2008; Kavzoglu et al. 2014; Zhang et al. 2023), artificial neural network (Yilmaz 2010; Wu et al. 2013), and random forest (Chen et al. 2018; Zhao et al. 2019) have oftentimes been adopted in the conventional landslide susceptibility modeling, and the study results indicate that the nonlinear relationships between various environmental factors (e.g., slope

angle, slope curvature, and geology) and the landslide occurrence probability in the spatial domain could be effectively captured by these machine learning techniques.

To realize an effective spatiotemporal prediction of rain-induced landslides, more and more attention is paid to coupling the triggering factor of rainfall with landslide susceptibility analyses. Some researchers (e.g., Segoni et al. 2018; Pradhan et al. 2019) proposed combining the landslide susceptibility map with rainfall thresholds using the matrix ensemble approach. The rainfall thresholds are often determined based on the statistical relationship between the landslide occurrence frequency and the corresponding rainfall records (Ko and Lo 2016; Rosi et al. 2016; Chen et al. 2017; Gao et al. 2018). As can be seen, the variation in the sensitivity of landslide occurrence to rainfall within the study area is not considered in these approaches. For example, two regions with different geological and topographical conditions could exhibit different sensitivities of landslide occurrence to rainfall (Jordanova et al. 2020; Wang et al. 2021a). To address this issue, new approaches have been proposed in recent studies (e.g., Wang et al. 2021b; Ng et al. 2021; Xiao et al. 2022), in which the maximum rolling  $x$ -h rainfall data (e.g., maximum rolling 12-h rainfall) is taken as a conditioning factor for the machine learning-based landslide susceptibility modeling. With the aid of these new approaches, the spatiotemporal probability of rain-induced landslides can be derived based on the fitted nonlinear relationships between landslide spatial susceptibility and maximum rolling  $x$ -h rainfall. It should be noted that the maximum rolling rainfall in a fixed time interval may not be able to describe the complex rainfall situation. For example, the maximum rolling 24-h rainfall data cannot be adopted to characterize the short-term heavy rainfall, whereas the maximum rolling 2-h rainfall data cannot reflect long-term small rainfall conditions. In other words, multiple rainfall data, not the maximum rolling rainfall in a fixed time interval, should be considered for spatiotemporal prediction of rain-induced landslides.

This study proposes an ensemble approach for spatiotemporal landslide susceptibility modeling, in which the dynamic rainfall index and the random forest method are ensembled; and, to consider complex rainfall conditions, multiple rainfall data (i.e., maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall) are integrated into a novel landslide conditioning factor, in terms of the maximum rolling rainfall index (MRRI). To further improve the accuracy of spatiotemporal landslide susceptibility modeling, a frequency ratio-based method (Liu et al. 2022; Zhang and Yan 2022) is adopted for selecting non-landslide samples in the study area. To depict the effectiveness and versatility of the proposed approach, landslide

susceptibility models are first developed based on the historical landslide data in the central area of Hong Kong. Then, the trained susceptibility models are applied to two historical rainstorms in Hong Kong. Based on the study results, the advantages and limitations of the proposed approach are discussed.

### Study area

The study area is situated in the central area of Hong Kong with an area of about 256 km<sup>2</sup> (113° 52' 53"–114° 5' 50" E, 22° 24' 27"–22° 33' 54" N), including Hong Kong Island and part of Kowloon and Lantau Island (Fig. 1a). Hong Kong is located at the mouth of the Pearl River Delta, South China, as shown in Fig. 1a. Noted that the natural mountains occupy about 60% of the land, many urban areas in Hong Kong are densely developed on hillsides. Hong Kong enjoys a subtropical monsoon climate characterized by a large amount of seasonal rainfall, which is mainly concentrated in the rainy season from June to September (AECOM and Lin 2015). Storms with high intensity and short duration are common in Hong Kong, and the hourly rainfall of some severe storms can exceed 200 mm (Gao et al. 2018). Most of the historical landslides reported in Hong Kong are shallow landslides, which are mainly triggered by frequent intense rainfall conditions on steep terrains; and the sliding materials are generally composed of saprolite, colluvium, and weathered rock (Lam et al. 2012).

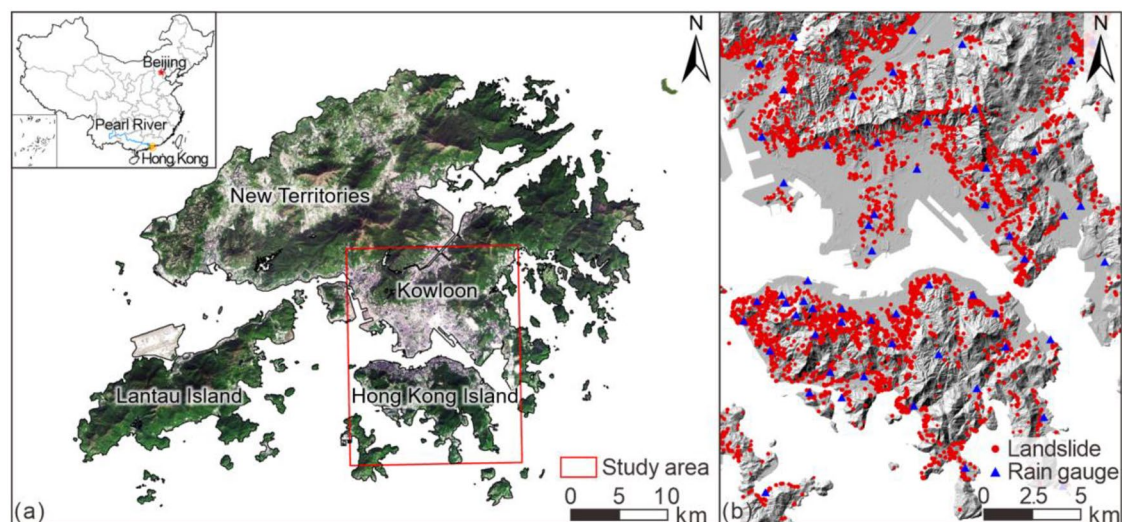
To ensure high-quality rainfall data acquisition, the Hong Kong Observatory (HKO) and Geotechnical Engineering Office (GEO) have installed plenty of automatic rain gauges in Hong Kong since the early 1980s, noted that 53 rain gauges are sparsely installed in the study area, and these rain gauges can provide real-time rainfall data at a 5-min interval. The spatial distribution of these gauges is shown in Fig. 1b. An inventory of historical landslides in Hong Kong known as the Natural Terrain Landslide Inventory (NTLI) was compiled by GEO in 1984 and has continuously enhanced based on aerial photographs and field investigations since then (Maunsell-Fugro Joint Venture 2007). A total of 4990 landslides were recorded in the study area from 1984 to 2009. The spatial distribution of historical landslides in the study area is illustrated in Fig. 1b.

### Spatiotemporal landslide susceptibility modeling approach

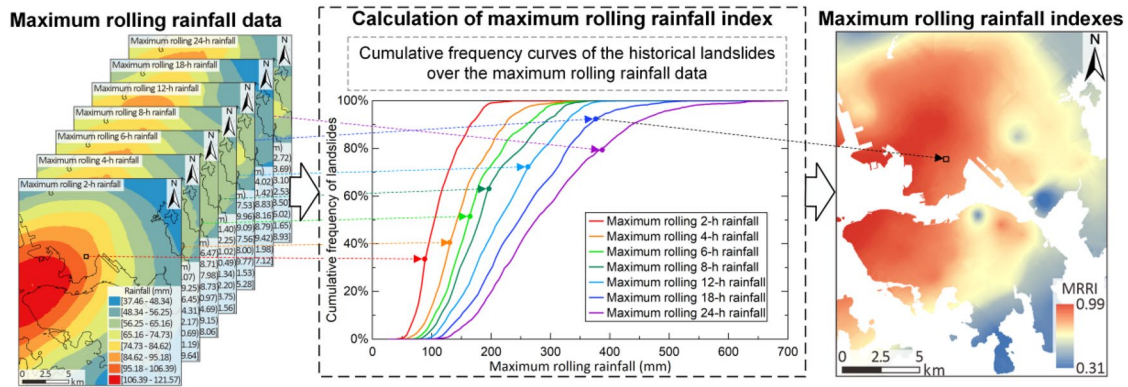
In this section, the formulation of MRRI, non-landslide sampling method, and random forest technique are briefly introduced; and the implementation procedures of the proposed spatiotemporal landslide susceptibility modeling approach are illustrated.

### Landslide triggering factor - maximum rolling rainfall index

The rolling  $x$ -h rainfall is defined as the rainfall recorded in  $x$  consecutive hours on a rain gauge (Chien-Yuan et al. 2008), whereas the maximum rolling  $x$ -h rainfall is defined as the maximum value of the rainfall in  $x$  consecutive hours on a rain gauge (Ng et al. 2021). To consider complex rainfall conditions, the MRRI which can be taken as a comprehensive index derived from the multiple rainfall data (i.e., maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall) is proposed herein. The MRRI is defined as the maximum cumulative frequency of historical landslides induced by the current maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data. Noted that this MRRI index is calculated from the maximum rolling rainfall recorded in all the durations of 2-, 4-, 6-, 8-, 12-, 18-, and 24-h. The procedures for calculating the MRRI index are summarized as follows (Fig. 2): (1) seven cumulative frequency curves of the historical landslides over the maximum rolling rainfall data are constructed for the durations of 2-, 4-, 6-, 8-, 12-, 18-, and 24-h, respectively; (2) seven cumulative frequency values are derived from the cumulative frequency curves constructed, based on the inputs of real-time maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data recorded at each position, respectively; and (3) the MRRI index is taken as the maximum cumulative frequency value derived in the previous step. Noted that in the historical landslide database constructed in the study area (i.e., 4,990 landslides), the precise occurrence timing could only be known for the limited number of historical landslides (i.e., 523 landslides). To ensure that there exist a sufficient number of historical landslides for constructing these seven cumulative frequency curves mentioned above, an assumption is made in this study: each landslide is assumed to be triggered



**Fig. 1** a Location of the study area. b Spatial distribution of historical landslides and rain gauges



**Fig. 2** Main procedure for calculating the maximum rolling rainfall index

by the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data in the year of landslide occurrence, respectively.

### Non-landslide sampling based on frequency ratio-based method

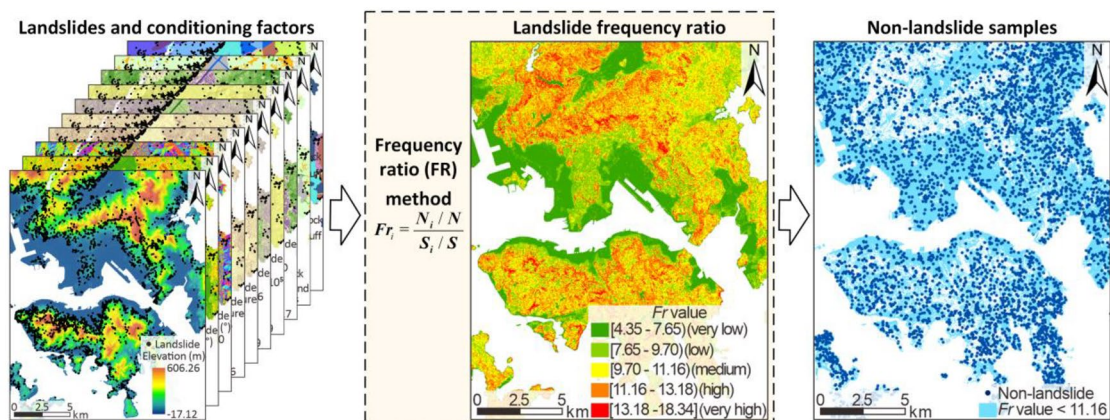
Supervised machine learning techniques are widely adopted in landslide susceptibility analyses (Chang et al. 2020). In general, the training dataset for building a supervised machine learning-based model includes landslide samples (i.e., historical landslides in the study area) and non-landslide samples. Noted that the selection of non-landslide data is vital for effective landslide susceptibility modeling, however, no effective criterion or rule has been established for the non-landslide sample selection. In some traditional analyses, non-landslide points are randomly selected in the study area, and then the non-landslide points that coincide with the landslide points are excluded. As can be seen, great uncertainty or error might exist in the non-landslide sample selection. To reduce the uncertainty or error in the non-landslide sample selection, a non-landslide sampling method based on frequency ratio (FR) is applied in this study, which can randomly generate non-landslide samples in areas with infrequent landslides. Noted that although this FR-based method is rough, it has been employed in some of the existing landslide

susceptibility modeling (Liu et al. 2022; Zhang and Yan 2022) and has been shown effective. The FR-based method is commonly employed to calculate the probabilistic relationship between dependent and independent variables (Ozdemir and Altural 2013). The index of FR is defined as the ratio of the landslide occurrence percentage to the area occupation percentage for various classes of every landslide conditioning factor in the study area, the formulation of which is provided below (Lee and Pradhan 2007).

$$Fr_i = \frac{N_i/N}{S_i/S} \quad (1)$$

where  $Fr_i$  denotes the frequency ratio of the  $i$ th class of a landslide conditioning factor;  $N_i$  is the number of landslides within the  $i$ th class of this factor; and  $N$  is the total number of landslides in the study area;  $S_i$  is the area of the  $i$ th class of this factor; and  $S$  is the total study area. A lower  $Fr_i$  value means that landslides are less likely to occur in the area.

The main procedures for generating non-landslide samples, with the FR-based method, are summarized in the following steps, as shown in Fig. 3. First, calculate the frequency ratio for each class of landslide conditioning factors. Second, aggregate the  $Fr$  values of



**Fig. 3** Main procedure for non-landslide sampling based on the FR-based method

all factors to obtain the overall  $Fr$  value for each position within the study area. Furthermore, divide the overall  $Fr$  values of the study area into five intervals by the Jenks optimization method (Jenks 1967), which are very low, low, medium, high, and very high. Finally, perform non-landslide random sampling in the areas with medium, low, or very low  $Fr$  values.

### Random forest technique

Random forest (RF), a machine learning technique proposed by Breiman (2001), can construct a multitude of decision trees to reveal the complex relationships between landslide occurrences and landslide conditioning factors (Catani et al. 2013). There are two stages involved in building a landslide susceptibility model using the RF technique: training and predicting stages. In the training stage, a large number of uncorrelated decision trees are generated with the bootstrap sample technique (Hastie et al. 2009). Noted that each tree is grown based on a random subset of input training samples, thus, each tree is unique. In the predicting stage, each tree yields assessment results of landslide susceptibility independently, and the final output is derived based on the unweighted majority of votes among all trees.

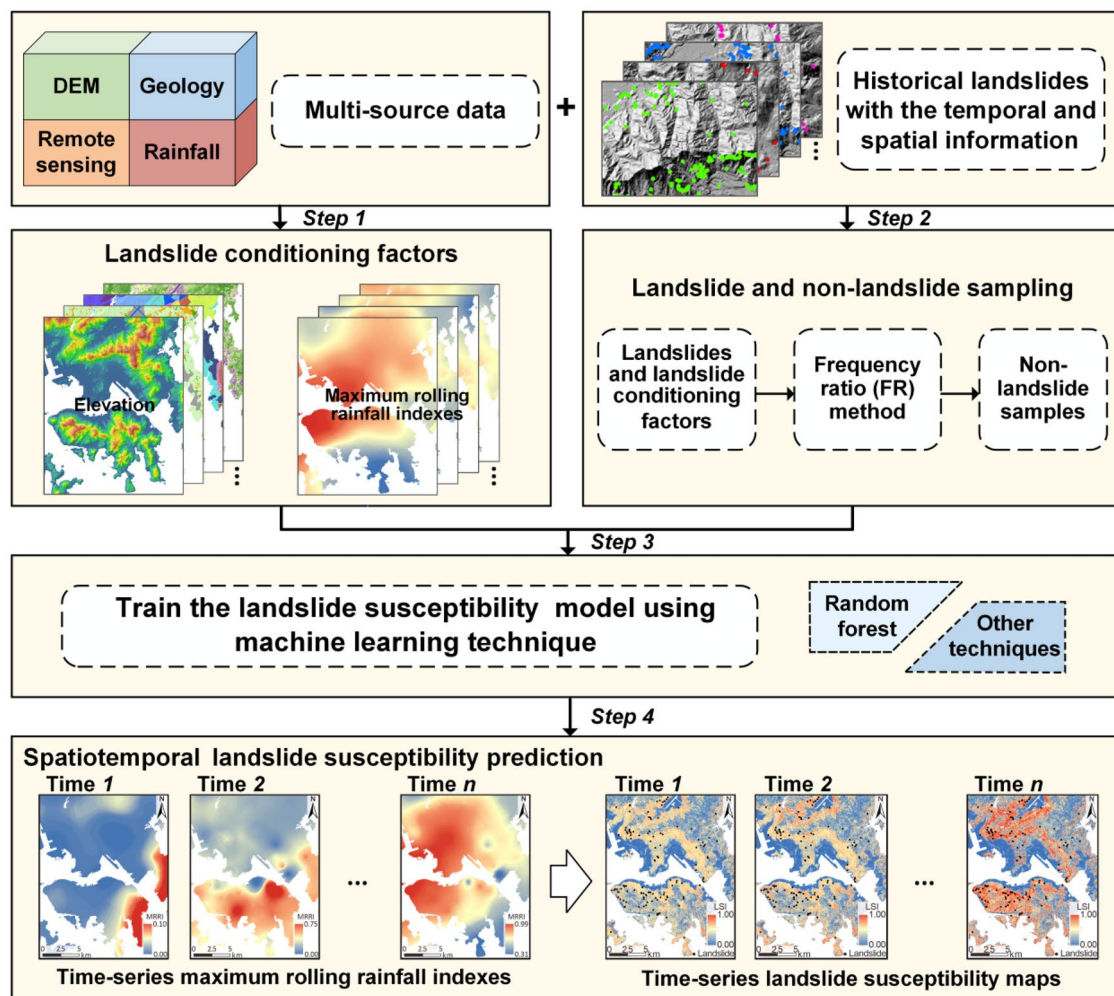
### Implementation procedures of the landslide susceptibility prediction approach

As shown in Fig. 4, the main procedures for implementing the proposed landslide susceptibility modeling approach are summarized in the following steps:

Step 1: Collect landslide conditioning factors, which could characterize the favorable environmental conditions for the development of landslides, from multi-source data such as digital elevation models, geological maps, and satellite images, and derive the MRRI indexes (i.e., landslide triggering factor) of historical landslides based on the historical rainfall and landslide information.

Step 2: Calculate  $Fr$  values in the whole study area from various landslide conditioning factors and historical landslides, and perform non-landslide random sampling in the study area based on computed  $Fr$  values, noted that the number of non-landslide samples should be equal to that of historical landslides.

Step 3: Extract the landslide conditioning factors and the condition triggering factor of the landslide and non-landslide samples, and train the landslide susceptibility model using the



**Fig. 4** Implementation procedures of the proposed spatiotemporal landslide susceptibility modeling approach

RF technique based on the information extracted, noted that the other advanced machine learning technique could also be adopted for training the landslide susceptibility model.

Step 4: Input time-series MRRRI indexes, obtained from the real-time maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data, to the trained landslide susceptibility model for predicting the spatial and temporal probabilities of landslide occurrence in the study area.

### Landslide susceptibility modeling in the study area

This section collects conditioning factors in the landslide susceptibility modeling in the study area from multi-source data. Ten landslide susceptibility models are then trained using the proposed landslide susceptibility modeling approach.

### Conditioning factors for landslide susceptibility modeling

The occurrence of landslides is closely related to the internal geological and external geomorphological conditions of the slope. Various environmental factors have been employed as landslide conditioning factors for landslide susceptibility modeling in existing studies, but these factors are selected on a case-by-case basis (e.g., Pradhan et al. 2019; Zhao et al. 2019; Gong et al. 2022). Based on previous studies on landslide susceptibility modeling in Hong Kong carried out by Yao et al. (2008) and Wang et al. (2021b), ten key conditioning factors, including elevation, slope angle, aspect, plan curvature, profile curvature, topographic wetness index (TWI), stream power index (SPI), normalized difference vegetation index (NDVI), bedrock, and surficial deposit, are chosen in this study (Fig. 5). The digital elevation model (DEM) of the study area is obtained from the high-precision LiDAR data of the entire Hong Kong acquired in January 2011, with a spatial resolution of 5.0 m/pixel. Most of the rain-induced landslides in Hong Kong are small-scale landslides, with a scar area ranging from 20 to 200 m<sup>2</sup> (Gao et al. 2021). Therefore, this DEM is considered to be sufficient to characterize the geomorphological conditions of natural terrain in the study area.

Five topographic conditioning factors (i.e., elevation, slope angle, aspect, plan curvature, profile curvature) can be directly derived from the DEM (Fig. 5a–e). Elevation is a basic terrain feature closely related to slope stability, which yields a considerable influence on the spatial distribution of conditioning factors such as land cover and rainfall. Slope angle is expected to be the most important geomorphological feature of the slope, as it is closely linked to the stress distribution within the terrain and partly reflects the accumulation of loose materials and rock weathering degree. Aspect is the compass direction that a slope faces and variation of slope aspect can lead to spatial differences in illumination and weathering, which may result in different soil textures, soil moisture, and vegetation development. Plan curvature and profile curvature are the terrain curvatures perpendicular and parallel to the direction of maximum slope, respectively, which could reflect the scene of flow across a surface.

Two hydrogeological conditioning factors (i.e., TWI and SPI) can be estimated from the DEM and the river network (Fig. 5f, g). Stream power index (SPI) measures the erosive power of the concentrated surface runoff, while topographic wetness index (TWI) indicates the percolation saturation of soil water (Moore et al. 1991). The formulations of SPI and TWI are provided below.

$$SPI = A_c \times \tan\phi \quad (2)$$

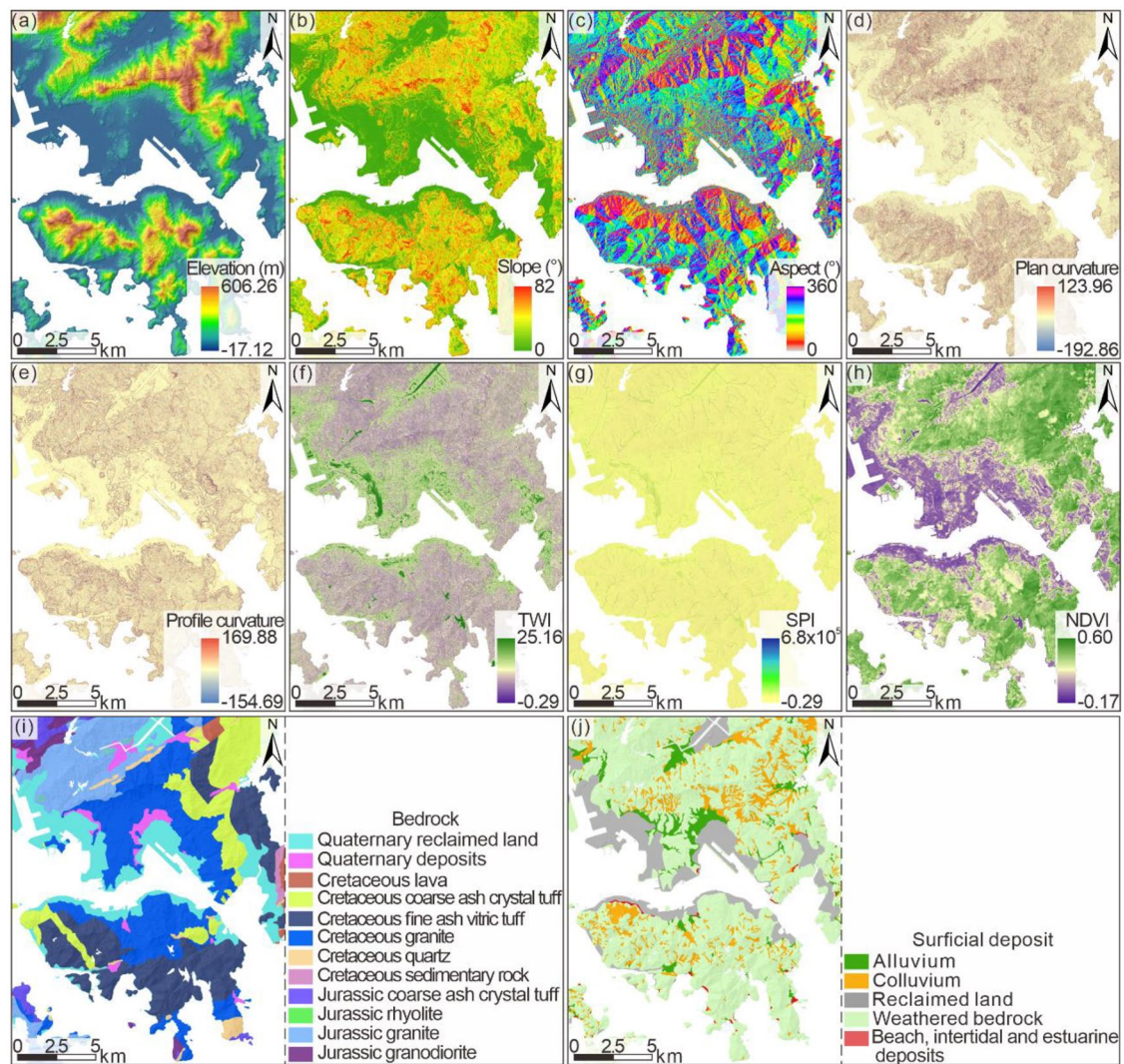
$$TWI = \ln\left(\frac{A_c}{\tan\phi}\right) \quad (3)$$

where  $A_c$  is the specific catchment area and  $\phi$  is the slope of the terrain.

Normalized difference vegetation index (NDVI), through which the degree of vegetation growth can be indicated, is usually adopted to reflect soil and hydrological conditions of the slope. Noted that soil and hydrological conditions exhibit significant influence on landslide occurrence. If the value of NDVI is equal to or smaller than 0, the related area is an area with non-vegetated features, such as water, bare ground (rock and soil), and artificial constructions, whereas areas with higher positive NDVI values are typically denser green vegetation regions (e.g., grass, shrub, and forest). It should be noted that landslides in the study area mainly occur in the rainy season (i.e., June to September), and the vegetation growth status is similar in the rainy season each year. As such, the temporal variation of the NDVI in the study area is not considered in the landslide susceptibility modeling conducted in this study. The NDVI, in this study, is derived from the satellite image of Landsat 8 with a resolution of 30 m/pixel, acquired on September 18, 2016. To be consistent with the other conditioning factors generated by the DEM, the resolution of the NDVI map is resampled to 5.0 m/pixel, as shown in Fig. 5h.

The maps of both bedrock and surficial deposits in Hong Kong, produced by the Hong Kong Geological Survey of the Civil Engineering and Development Department (CEDD), have been open to the public since 2006. The geological maps of the study area are converted into raster format with a resolution of 5.0 m/pixel. In total, 12 types of bedrock are distributed in the study area (Fig. 5i), which could be divided into three geological periods (i.e., Jurassic, Cretaceous, Quaternary). Granitic rock (i.e., Jurassic granitic and Cretaceous granitic) and volcanic rock (i.e., Jurassic coarse ash, Cretaceous lava, fine ash vitric tuff and crystal tuff) are the dominant bedrocks, which cover 38.9% and 25.5% of the land area, respectively. The surficial deposits in the study area could be divided into alluvium, colluvium, reclaimed land, weathered bedrock, and beach, intertidal, and estuarine deposits (Fig. 5j). Because of the warm and humid environment, the rate of bedrock weathering in the study area is high and weathered mantles are widely distributed on the slopes.

As mentioned above, only 53 rain gauges are sparsely distributed in the study area, and the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data at positions without rain gauges are interpolated from the rainfall data recorded by these gauges, while the spatial unit adopted is 5.0 × 5.0 m. In this study, the ordinary Kriging interpolation method in the Geostatistical Analyst module of ArcGIS software is adopted, and the exponential model (Leung and Law 2002) is adopted for building the semi-variogram while the separation distance of the semi-variograms established is about 15.0 km. Meanwhile, at least three adjacent rain gauges are utilized for interpolating the rainfall data at each position. Based on the historical rainfall and landslide information, seven cumulative frequency curves of the historical landslides over the maximum rolling rainfall data are constructed for the durations of 2-, 4-, 6-, 8-,



**Fig. 5** Landslide conditioning factors of the study area: **a** elevation; **b** slope; **c** aspect; **d** plan curvature; **e** profile curvature; **f** topographic wetness index (TWI); **g** stream power index (SPI); **h** normalized difference vegetation index (NDVI); **i** bedrock; **j** surface deposit

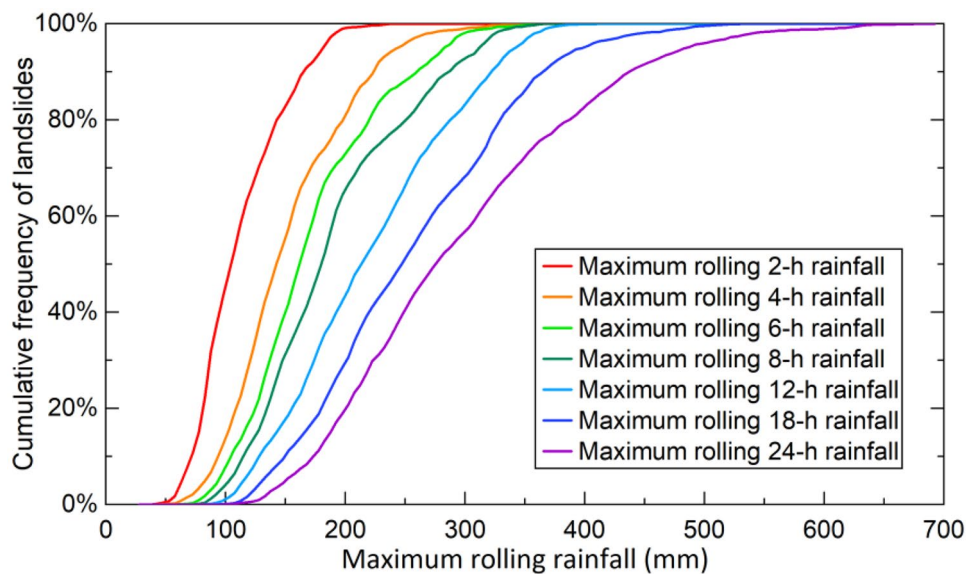
12-, 18-, and 24-h, respectively, as illustrated in Fig. 6. MRRI indexes of historical landslides that is adopted for training the landslide susceptibility model are derived according to these cumulative frequency curves. In contrast, the MRRI indexes adopted for the landslide susceptibility prediction are estimated based on the real-time rainfall data and the cumulative frequency curves shown in Fig. 6.

### Data preprocessing and landslide susceptibility modeling

As mentioned above, a total of 4990 landslides are recorded in the study area from 1984 to 2009; among which, 4764 historical landslides are selected as the landslide samples for the landslide susceptibility modeling, noted that 226 landslides induced by the rainfall on August 19–20, 2005, and June 6–7, 2008, are retained to later demonstrate the application of the trained landslide susceptibility model. To generate the non-landslide samples in the study area, the FR-based method is adopted. The  $Fr$  values of various

classes of ten landslide conditioning factors are calculated, as detailed in Table 1. The map of overall  $Fr$  values in the study area is further obtained by aggregating the  $Fr$  values of all factors and divided into five zones (Fig. 7a): very low, low, medium, high, and very high. Finally, 4764 non-landslide samples are randomly generated in the areas with medium, low, or very low  $Fr$  values (Fig. 7b).

It is known that unimportant landslide conditioning factors and multicollinearity among the conditioning factors could degrade the accuracy of the landslide susceptibility models trained (Zhou et al. 2018). In this study, the variance inflation factor (VIF) and tolerance indexes are computed and adopted to identify the multicollinearity among the conditioning factors, and the information gain ratio is adopted to assess the relative importance of each conditioning factor. The VIF is an indicator that is often estimated to measure the degree of variance of a regression coefficient induced by the multicollinearity among input variables (O'Brien 2007). Tolerance is the reciprocal of



**Fig. 6** Cumulative frequency curves of the historical landslides over the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data in the study area

VIF. When the value of the estimated VIF is greater than 5.0 or the computed tolerance is less than 0.2, multicollinearity exists among the input variables. The information gain ratio is a normalized information gain based on the entropy of values in each class, the mathematical formulation of which can be available in Ghasemi et al. (2020). Generally, a conditioning factor with a higher value of the information gain ratio indicates more importance of this factor in landslide susceptibility modeling. In addition, a random variable varying from 0 to 1.0 is sampled to assess its relative importance in landslide susceptibility modeling. The information gain ratio of this random variable could be taken as a benchmark to identify the unimportant landslide conditioning factors (Luti et al. 2020; Segoni et al. 2020).

As tabulated in Table 2, the smallest tolerance and the highest VIF of the landslide conditioning factors are 0.452 and 2.215, respectively; thus, no multicollinearity is found among the conditioning factors in this study. Figure 8 shows the relative importance of eleven landslide conditioning factors and the random variable. Because the random variable is not related to the landslide occurrence, its information gain ratio is the lowest as expected. The information gain ratio of each conditioning factor is greater than that of the random variable, indicating that all the selected conditioning factors contribute to the development of the landslide susceptibility model. Specifically, the slope angle yields the highest information gain ratio (i.e., 0.131), followed by the profile curvature (i.e., 0.130). The information gain ratio of the MRRI is relatively high (i.e., 0.057), indicating that rainfall has a strong effect on the landslide occurrence in the study area. In comparison to the bedrock factor (i.e., 0.037), the surficial deposit factor (i.e., 0.041) appears to be more closely related to the occurrence of landslides, which is consistent with the engineering experience (Ko and Lo 2016; Gao et al. 2018). Noted that although the aspect yields the relatively low information gain ratio (i.e., 0.007), excluding this conditioning factor could degrade the effectiveness of the landslide susceptibility modeling. Thus, our landslide

susceptibility modeling keeps this least informative conditioning factor. In summary, all eleven conditioning factors are vital for the landslide susceptibility modeling in the study area.

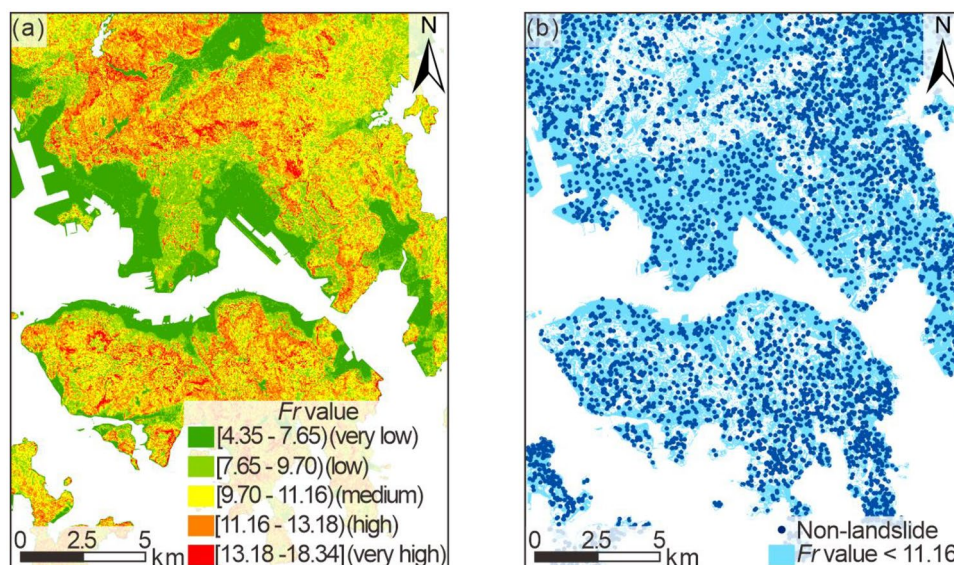
In the landslide susceptibility modeling with the RF technique, the information of landslide and non-landslide samples with all conditioning factors is taken as the input, while the landslide susceptibility index (i.e., landslide 1, non-landslide 0) is taken as the output. For ease of the training of the landslide susceptibility model, the categorical conditioning factors (i.e., bedrock and surficial deposit) are assigned numerical labels, and continuous conditioning factors (i.e., elevation, slope angle, aspect, plan curvature, profile curvature, TWI, SPI, NDVI, and MRRI) are normalized using the min–max normalization method (Salehpour et al. 2021). Among the 4764 landslide samples and 4764 non-landslide samples, 80% are randomly selected for the model training, whereas the remaining samples are employed to validate the trained landslide susceptibility model. The number of decision trees and the maximum depth of the tree are 500 and 5, respectively. Two indicators, in terms of the accuracy and receiver operating characteristic (ROC) curve, are adopted to assess the performance of the landslide susceptibility model trained. The accuracy is the proportion of the correct predictions in the total testing samples. In the ROC curve, the false positive and true positive rates are plotted on the X and Y axes, respectively. The area under the curve (AUC) is adopted to measure the probability of correct classification. The performance of a landslide susceptibility model with accuracy and AUC close to 1.0 is considered good. To avoid the randomness induced in the landslide susceptibility model training, the landslide susceptibility modeling is conducted ten times (with the proposed approach); as such, ten susceptibility models can be derived. The average accuracy and AUC of the testing samples are 86.30% and 0.933 (Fig. 9), which demonstrates the effectiveness of the trained landslide susceptibility models. The plots in Fig. 9 also suggest that the variability of the trained landslide susceptibility models induced by

**Table 1** *Fr* values of landslide conditioning factors

Landslide conditioning factor	Category	Number of landslides $N_i$	Area $S_i$ (km <sup>2</sup> )	$Fr$	Landslide conditioning factor	Category	Number of landslides $N_i$	Area $S_i$ (km <sup>2</sup> )	$Fr$
Aspect	N	584	27.97	1.08	TWI	[-0.29, 2.00)	245	4.90	2.59
	NE	473	30.90	0.79		[2.00, 3.00)	1222	39.43	1.61
	E	517	33.69	0.80		[3.00, 4.00)	1559	62.94	1.28
	SE	599	32.56	0.95		[4.00, 5.00)	985	48.90	1.04
	S	647	33.62	1.00		[5.00, 6.00)	468	33.63	0.72
	SW	841	36.78	1.19		[6.00, 7.00)	246	24.59	0.52
	W	746	34.68	1.12		[7.00, 8.00)	99	15.59	0.33
	NW	583	28.65	1.06		[8.00, 25.16]	166	28.87	0.30
Slope (°)	[0, 15)	530	108.93	0.25	Bedrock	$Q_r$	42	38.67	0.06
	[15, 20)	332	22.31	0.77		$Q_d$	52	7.79	0.35
	[20, 25)	531	28.15	0.98		$C_l$	17	3.18	0.28
	[25, 30)	796	34.37	1.20		$C_c$	728	34.34	1.10
	[30, 35)	914	32.64	1.45		$C_f$	1012	54.96	0.96
	[35, 40)	793	19.51	2.11		$C_g$	1772	66.23	1.39
	[40, 82]	1094	12.94	4.39		$C_q$	56	4.79	0.61
Elevation (m)	[17.12, 100.00)	2916	140.27	1.08	Surficial deposit	$C_s$	19	1.49	0.66
	[100.00, 200.00)	1345	53.61	1.30		$J_c$	107	7.87	0.71
	[200.00, 300.00)	426	38.16	0.58		$J_r$	17	0.22	3.99
	[300.00, 400.00)	204	19.12	0.55		$J_{g1}$	1044	34.44	1.57
	[400.00, 500.00)	90	6.35	0.74		$J_{g2}$	124	4.86	1.32
	[500.00, 606.26]	9	1.33	0.35		$A_l$	126	13.88	0.46
Plan curvature	[-192.86, -0.05)	2666	112.58	1.23		$C_o$	580	26.47	1.14
	[-0.05, 0.05)	96	18.03	0.28		$R_l$	57	35.05	0.08
	[0.05, 123.96]	2228	128.23	0.90		$W_b$	14	1.07	0.68
Profile curvature	[-154.69, -0.05)	2102	121.20	0.90	SPI	$B_{ie}$	4213	182.37	1.20
	[-0.05, 0.05)	69	15.87	0.23		[-0.29, 2.00)	2402	173.17	0.72
	[0.05, 169.88]	2819	121.77	1.20		[2.00, 4.00)	918	29.66	1.61
NDVI	[-0.17, 0.20)	786	73.25	0.56		[4.00, 8.00)	696	21.83	1.65
	[0.20, 0.49)	4137	177.40	1.21		[8.00, $6.80 \times 10^5$ ]	974	34.18	1.48
	[0.50, 0.60]	67	8.19	0.42					

NDVI normalized difference vegetation index, TWI topographic wetness index, SPI stream power index,  $Q_r$  Quaternary reclaimed land,  $Q_d$  Quaternary deposits,  $C_l$  Cretaceous lava,  $C_c$  Cretaceous coarse ash crystal tuff,  $C_f$  Cretaceous fine ash vitric rock,  $C_g$  Cretaceous granite,  $C_q$  Cretaceous quartz,  $C_s$  Cretaceous sedimentary rock,  $J_c$  Jurassic coarse ash crystal tuff,  $J_r$  Jurassic rhyolite,  $J_{g1}$  Jurassic granite,  $J_{g2}$  Jurassic granodiorite,  $A_l$  alluvium,  $C_o$  colluvium,  $R_l$  reclaimed land,  $W_b$  weathered bedrock,  $B_{ie}$  beach, intertidal, and estuarine deposits





**Fig. 7** Non-landslide sampling in the study area. **a** The map of overall  $Fr$  values for the study area. **b** Distribution of the non-landslide samples

training randomness is small. For illustration purposes, only the averaging results of the ten susceptibility models are presented in the following text.

#### Application of the trained landslide susceptibility models to two rainfall events

Two historical rainfall events are selected to demonstrate the spatiotemporal predictive ability of the established landslide susceptibility models. The first rainfall event occurred on August 19–20, 2005, a moderate rainfall event with a long duration that triggered

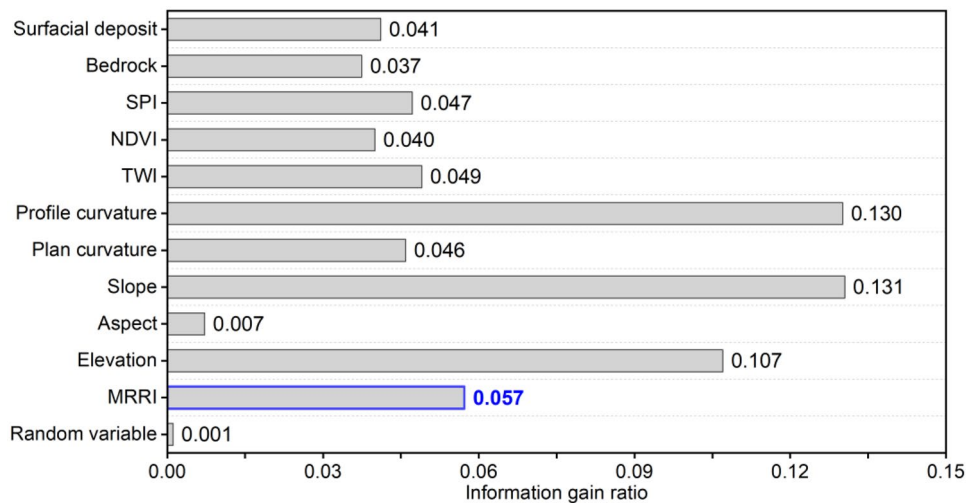
93 landslides in the study area. The other rainfall event occurred on June 6–7, 2008, which was one of the most severe rainstorms recorded in Hong Kong, and this rainfall event triggered 133 landslides in the study area. Figure 10 shows the spatial distributions of the MRRI index in the study area during the two rainfall events.

In the first rainfall event that occurred in August 2005, the peak center of the MRRI gradually moves from the southwestern Kowloon to the central Kowloon (Fig. 10a–h). The relatively quick increase in the MRRI index in the period from 21:00 on August 19 to 00:00 on August 20, 2005, and that from 09:00 to 12:00 on August 20, 2005, confirms the increase of the short-duration rainfall intensity, while the slow increase of the MRRI index in the remaining periods indicates the weak rainfall intensity. Compared to the first rainfall event, the spatial variation of the MRRI index in the second rainfall event that occurred in June 2008 is much more complicated. For example, in the early stage of the second rainfall event (Fig. 10i), the maximum MRRI index is only 0.10, and the peak center is located in the southeastern part of Hong Kong Island, indicating that the rainfall intensity is small. Then, the MRRI index increases slowly throughout the study area (Fig. 10j–n). The MRRI index in the entire study area increases rapidly from the southeastern after 6:00 on June 7, 2008. All exceed 0.30 within 6 h, especially in the western region, with a maximum value of 0.99 (Fig. 10o, p). As can be seen, the characteristics of these two rainfall events are different: widespread moderate rainfall prevails in the first rainfall event, while more intense short-duration rainfall hits the entire study area in the second rainfall event. It is evident from this that the MRRI index's spatial distribution can well capture the rainfall characteristics.

On the basis of the spatial distributions of the MRRI index derived in these two rainfall events, the spatiotemporal landslide susceptibility maps are readily obtained with the trained landslide susceptibility models. For ease of landslide susceptibility mapping, the study area is divided into 10,268,219 grids with a size of  $5.0 \times 5.0$  m. Shown in Fig. 11 are the spatial distributions of the landslide susceptibility

**Table 2** Variance inflation factors and tolerances of landslide conditioning factors

Landslide conditioning factor	Variance inflation factor	Tolerance
Elevation	1.235	0.809
Slope	1.927	0.519
Aspect	1.033	0.968
Plan curvature	1.454	0.688
Profile curvature	1.250	0.800
Bedrock	1.126	0.888
Surficial deposit	1.137	0.879
Topographic wetness index	2.215	0.452
Stream power index	1.140	0.877
Normalized difference vegetation index	1.352	0.740
Maximum rolling rainfall index	1.022	0.979



**Fig. 8** Information gain ratios of landslide conditioning factors and random variable

index (LSI) obtained in these two rainfall events. As can be seen in Fig. 11, in the early stage of the first rainfall event (i.e., from 21:00 on August 19 to 00:00 on August 20, 2005), the areas with high LSI values (i.e.,  $LSI > 0.8$ ) are mainly located in southwestern Kowloon and northern Hong Kong Island (Fig. 11a–d), then the areas with high LSI values gradually distribute in the whole mountainous areas of the study area (Fig. 11e–h). In the early stage of the second rainfall event, the LSI value in the whole study area is not high (Fig. 11i); in the middle stage (i.e., from 21:00 on June 6 to 3:00 on June 7, 2008), the areas with high LSI values are mainly located in northern Kowloon and southeastern Hong Kong Island (Fig. 11j–n); and finally, the areas with high LSI values expand to the entire mountainous areas (Fig. 11o, p), due to the rapid increase of the overall MRRRI index. In addition, the locations of all real rain-induced landslides in the two rainfall

events are marked in Fig. 11. Note that only the occurrence dates of these landslides could be available while the precise occurrence timing cannot be available. Thus, maps showing the dynamic evolution of real rain-induced landslides versus time are not provided in this study. The time-series landslide susceptibility maps shown in Fig. 11 depict that the expansion trend of the high LSI area corresponds well to the spatial distribution of the real landslides. Thus, the effectiveness and versatility of the trained landslide susceptibility models are demonstrated.

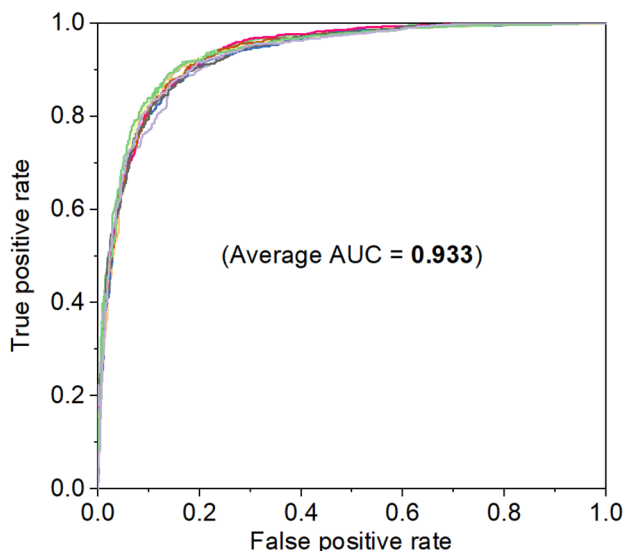
To further illustrate the effectiveness of the trained landslide susceptibility models, the percentage of the areas with high LSI values (i.e.,  $LSI > 0.8$ ), the number of real landslides located in the high LSI areas, and the average hourly rainfall in the study area during these two rainfall events are derived, and the results are plotted in Fig. 12. As can be seen, the variation trend of the percentage of the areas with high LSI values matches that of the number of real landslides. Furthermore, the areas with high LSI values increase rapidly from 6:00 to 11:00 on June 7, 2008, due to the short-duration heavy rainfall in the second rainfall event. In contrast, due to the moderate rainfall pattern, the areas with high LSI values increase relatively slowly in the first rainfall event. The plots in Fig. 12 also demonstrate that the variation of the percentage of the areas with high LSI values corresponds well with the rainfall condition. Thus, complex rainfall conditions could be effectively captured by the trained landslide susceptibility models.

### Discussion

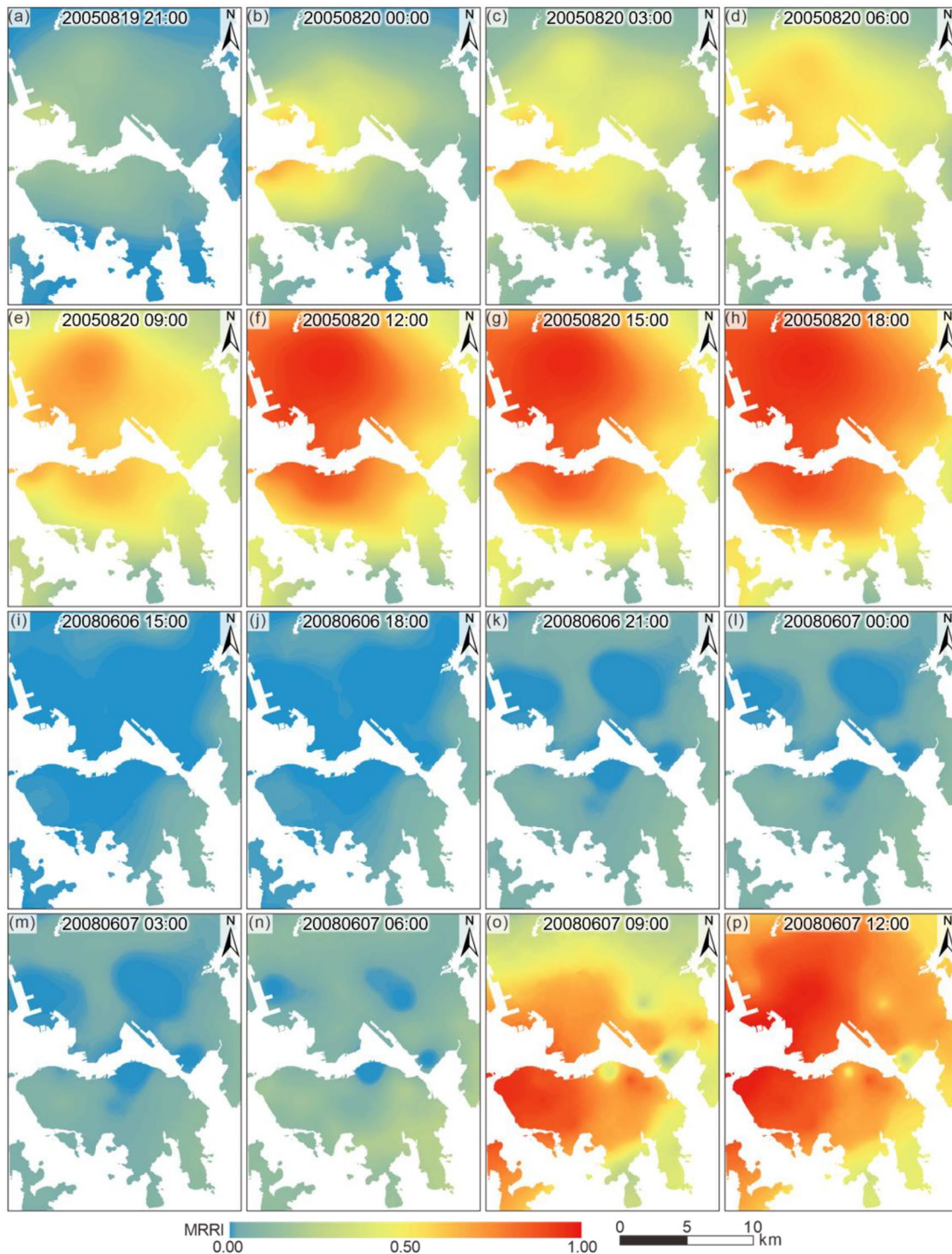
In this section, comparative analyses are first conducted to demonstrate the advantages of the proposed landslide susceptibility modeling approach over some existing methods; then, the limitations of the proposed approach are discussed.

#### Comparisons between the proposed approach and the existing methods

One of the important components of the proposed approach is the use of the MRRRI as a landslide triggering factor to capture the rainfall conditions. In contrast, the existing landslide susceptibility



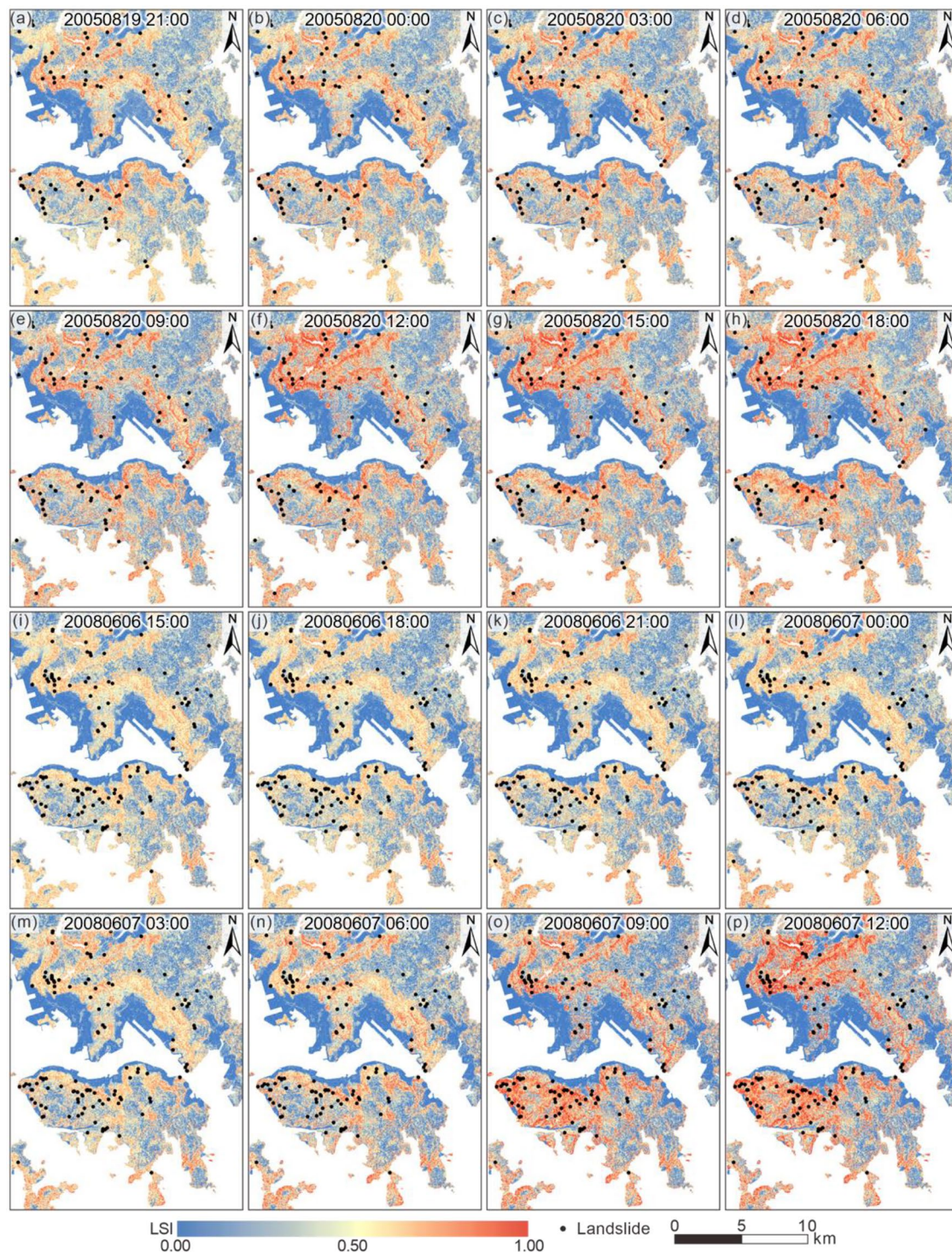
**Fig. 9** Receiver operating characteristic (ROC) curves of the landslide susceptibility models trained by the proposed approach



**Fig. 10** Spatial distributions of MRRRI index in the study area during the two rainfall events: **a–h** from 21:00 on August 19 to 18:00 on August 20, 2005; **i–p** from 15:00 on June 6 to 12:00 June 7, 2008

studies frequently employ the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data (e.g., Wang et al. 2021b; Ng et al. 2021; Xiao et al. 2022). The relative importance of the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall as the triggering factor for landslide susceptibility modeling in the study area is quantitatively assessed

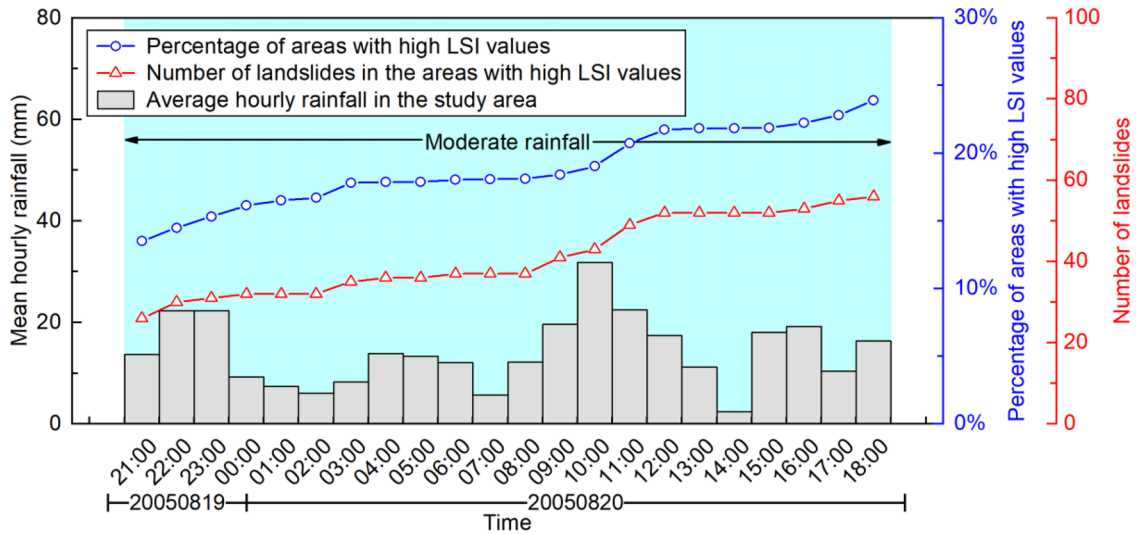
with the information gain ratio (Fig. 13). Noted that the information gain ratios of all the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall factors are smaller than that of the MRRRI (i.e., 0.057), showing that the MRRRI contributes more to susceptibility modeling. As the maximum rolling 24-h rainfall factor yields the highest



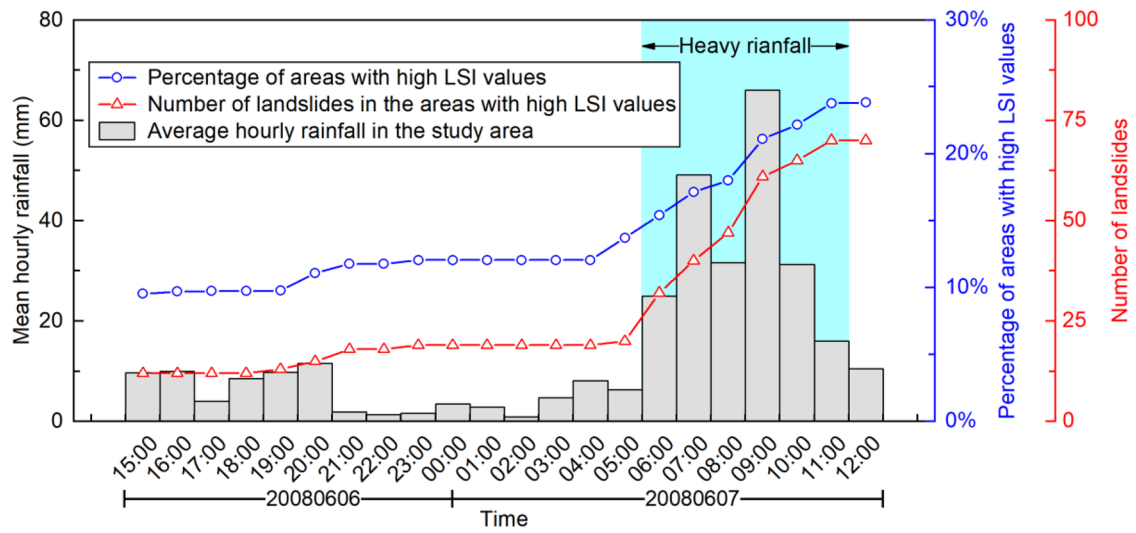
**Fig. 11** Spatial distributions of the LSI index in the study area during the two rainfall events: **a–h** from 21:00 on August 20, 2005; **i–p** from 15:00 on June 6 to 12:00 June 7, 2008

information gain ratio among the conventional seven maximum rolling rainfall factors, it is then selected to implement the conventional landslide susceptibility modeling in the further comparative analysis, noted that the same training and testing samples, as those adopted in "Landslide susceptibility modeling in the study area"

section, are adopted in the conventional landslide susceptibility modeling. The susceptibility modeling is also conducted ten times using the maximum 24-h rolling rainfall, and the average accuracy and AUC of the trained landslide susceptibility models are 81.44% and 0.898 (Fig. 14a), noted that the average accuracy and AUC of



(a)



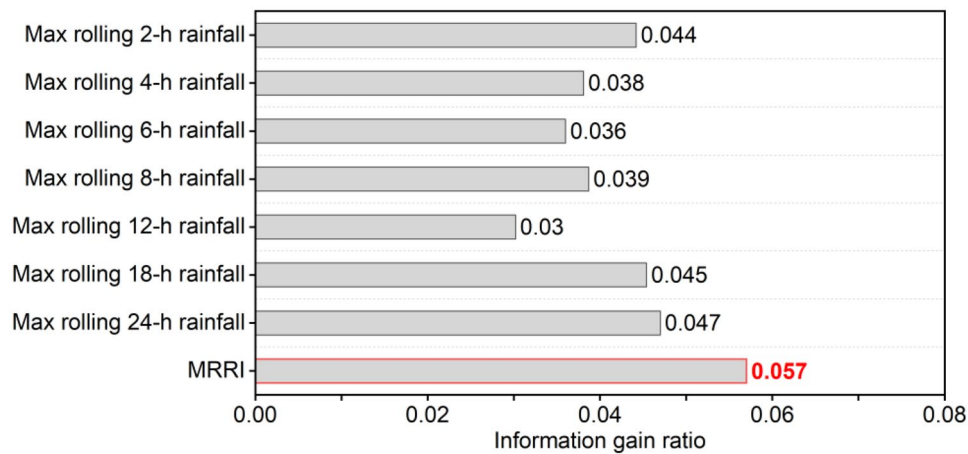
(b)

**Fig. 12** Percentage of the areas with high LSI values (i.e., >0.8), the number of real landslides located in the high LSI areas, and the average hourly rainfall in the study area during the two rainfall events: **a** from 21:00 on August 19 to 18:00 on August 20, 2005; **b** from 15:00 on June 6 to 12:00 June 7, 2008

the landslide susceptibility models trained by the maximum 24-h rolling rainfall are smaller than those trained by the MRRI (i.e., average accuracy, 86.30%; average AUC, 0.933). Thus, the MRRI is shown more effective for landslide susceptibility modeling, in comparison to the maximum rolling 24-h rainfall factor adopted in the existing landslide susceptibility modeling.

The other vital component of the landslide susceptibility modeling approach is the non-landslide sampling based on the FR-based method. In contrast, the non-landslide samples in the conventional landslide susceptibility modeling are randomly selected in the study area (e.g., Chen et al. 2018; Zhao et al. 2019). Thus, a comparative analysis is further conducted to illustrate the effectiveness of the FR-based method. In this comparative

analysis, 4764 non-landslide samples are randomly selected in the study area. The landslide susceptibility modeling is then conducted ten times using the non-landslide samples that are randomly selected in the study area, and the average accuracy and AUC of the trained landslide susceptibility models are 82.25% and 0.903 (Fig. 14b), noted that the average accuracy and AUC of the landslide susceptibility models trained by the randomly selected non-landslide samples are smaller than those trained by the non-landslide samples obtained with the FR-based method (i.e., average accuracy, 86.30%; average AUC, 0.933). Thus, the non-landslide samples obtained with the FR-based method are more effective for landslide susceptibility modeling, in comparison to the randomly selected non-landslide samples. However,



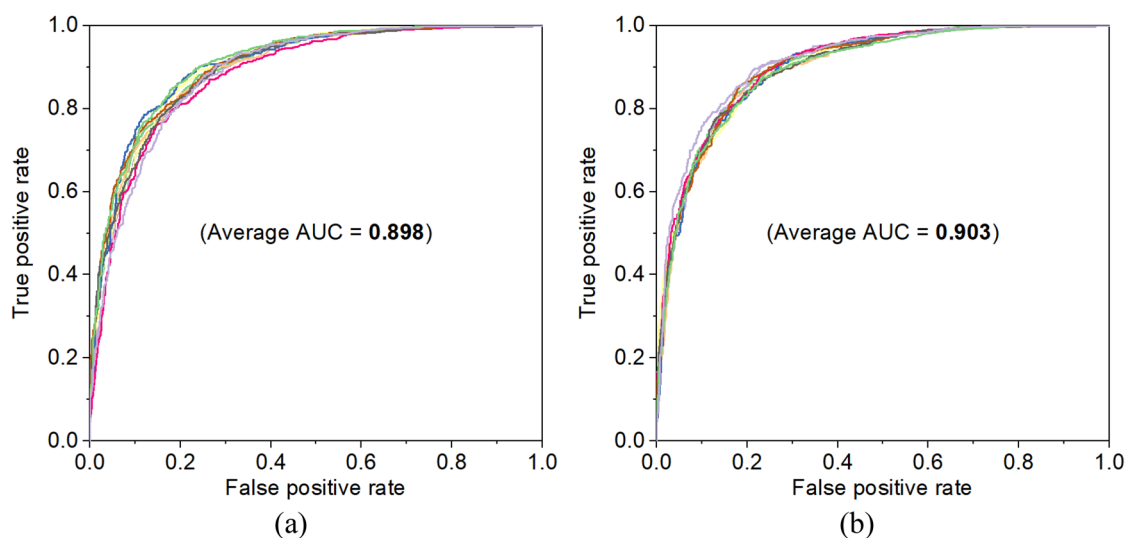
**Fig. 13** Information gain ratios of maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall and MRRI

it should be noted that although the non-landslide sampling strategy based on the FR-based method is shown to be effective in this study, the non-landslide samples, within the context of this FR-based method, are indeed generated from the landslide samples through data analyses (not from field surveys), and the data analyses might be tricky. Thus, the method for obtaining non-landslide samples merits further investigation and improvement.

#### Limitations of the proposed landslide susceptibility modeling approach

Noted that although the proposed landslide susceptibility modeling approach is shown in this study area and the superiorities over the existing susceptibility modeling methods are demonstrated, the proposed landslide susceptibility modeling approach

is not perfect and the following aspects warrant further investigation: (1) although the complex and dynamic rainfall conditions can be considered in the proposed method, the real-time spatiotemporal prediction of landslides under a rainfall event could not be reached, as the temporal delays between the landslide occurrence and rainfall are not included; (2) the output of the landslide susceptibility prediction is the landslide susceptibility index (LSI) ranging from 0 to 1.0, this index is more like a landslide occurrence probability, not a deterministic value, implying that the landslide susceptibility modeling results could only be interpreted in a probabilistic manner; and (3) the proposed approach is a data-driven statistical approach and the modeling results are strongly affected by the data quality and quantity; as the physical mechanism of the landslide cannot be included, the proposed approach cannot take the advantage of the accumulated knowledge of landslide mechanism.



**Fig. 14** Receiver operating characteristic (ROC) curves of the landslide susceptibility models: **a** the models trained by the maximum 24-h rolling rainfall; **b** the models trained by the randomly selected non-landslide samples

## Conclusions

This study proposed an ensemble approach for spatiotemporal landslide susceptibility modeling, in which the dynamic rainfall index and the random forest method are ensembled. Within the context of the proposed approach, the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall data are integrated into the maximum rolling rainfall index (MRRI) to consider complex rainfall conditions; and the frequency ratio (FR)-based method is taken for selecting non-landslide samples in the proposed approach. To illustrate the effectiveness and versatility of the proposed approach, ten landslide susceptibility models were first trained based on the historical landslide data collected in the central area of Hong Kong from 1984 to 2009; the trained susceptibility models were then applied to generate the spatiotemporal prediction of landslides under two historical rainfall events. The following conclusions are reached based upon the results presented.

1. The spatiotemporal landslide susceptibility modeling approach proposed was shown effective in the studies conducted. The time-series landslide susceptibility maps derived from the landslide susceptibility models trained by the proposed approach were in good agreement with the spatial distribution of real landslides under two rainfall events.
2. Compared to the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall adopted in the conventional landslide susceptibility modeling, the MRRI index proposed in this study was more effective. The information gain ratios of all the maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall were smaller than that of the MRRI; thus, the MRRI contributes more to landslide susceptibility modeling. The results of comparative analyses suggested that the landslide susceptibility models trained by the MRRI could perform better than those trained by the traditional maximum rolling 2-, 4-, 6-, 8-, 12-, 18-, and 24-h rainfall. Furthermore, complex rainfall conditions could be well captured by the MRRI index formulated.
3. Comparative analyses conducted also confirmed the effectiveness of the FR-based method for non-landslide sampling in the landslide susceptibility modeling. The non-landslide samples obtained with the FR-based method were more effective in the landslide susceptibility modeling, in comparison to the non-landslide samples that were randomly selected.

## Acknowledgements

The financial support is gratefully acknowledged. Thanks are also extended to the editor and the two anonymous reviewers who handled this paper during the reviewing process.

## Author contribution

Conceptualization, T. Ren and L. Gao; methodology, T. Ren and W. Gong; writing—original draft preparation, T. Ren; writing—review and editing, L. Gao and W. Gong; funding acquisition, L. Gao and W. Gong. All authors have read and agreed to the published version of the manuscript.

## Funding

The research work was funded by National Natural Science Foundation of China (52209118), the Science and Technology Development Fund, Macau SAR (File no.: SKL-IOTSC(UM)-2021-2023, 0030/2020/A1, 0029/2022/A1), UM Research Grant (File no.: MYRG2020-00072-IOTSC, MYRG2022-00090-IOTSC), Shenzhen Science and Technology Innovation Committee (SGDX20210823103805043), CORE, and the Outstanding Youth Foundation of Hubei Province, China (No. 2022CFA102). CORE is a joint research center for ocean research between QNLM and HKUST.

## Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- AECOM, Lin B (2015) 24-hour probable maximum precipitation updating study. GEO Report No. 314. Geotechnical Engineering Office, Hong Kong Special Administration Region. [https://www.cedd.gov.hk/eng/publications/geo/geo-reports/geo\\_rpt314/index.html](https://www.cedd.gov.hk/eng/publications/geo/geo-reports/geo_rpt314/index.html)
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazards Earth Syst Sci* 13(11):2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Chang Z, Du Z, Zhang F, Huang F, Chen J, Li W, Guo Z (2020) Landslide susceptibility prediction based on remote sensing images and GIS: comparisons of supervised and unsupervised machine learning models. *Remote Sens* 12(3):502. <https://doi.org/10.3390/rs12030502>
- Chen CW, Oguchi T, Hayakawa YS, Saito H, Chen H (2017) Relationship between landslide size and rainfall conditions in Taiwan. *Landslides* 14(3):1235–1240. <https://doi.org/10.1007/s10346-016-0790-7>
- Chen W, Xie X, Peng J, Shahabi H, Hong H, Bui DT, Duan Z, Li S, Zhu AX (2018) GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA* 164:135–149. <https://doi.org/10.1016/j.catena.2018.01.012>
- Cheng Z, Gong W, Tang H, Juang CH, Deng Q, Chen J, Ye X (2021) UAV photogrammetry-based remote sensing and preliminary assessment of the behavior of a landslide in Guizhou. *China Eng Geol* 289:106172. <https://doi.org/10.1016/j.enggeo.2021.10617>
- Chien-Yuan C, Lien-Kuang C, Fan-Chieh Y, Sheng-Chi L, Yu-Ching L, Chou-Lung L, Yu-Ting W, Kei-Wai C (2008) Characteristics analysis for the flash flood-induced debris flows. *Nat Hazards* 47:245–261. <https://doi.org/10.1007/s11069-008-9217-7>
- Dai FC, Lee CF (2001) Terrain-based mapping of landslide susceptibility using a geographical information system: a case study. *Can Geotech J* 38(5):911–923. <https://doi.org/10.1139/t01-021>
- Gao L, Zhang LM, Cheung RWM (2018) Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. *Landslides* 15(4):727–740. <https://doi.org/10.1007/s10346-017-0904-x>

- Gao L, Zhang LM, Chen HX, Fei K, Hong Y (2021) Topography and geology effects on travel distances of natural terrain landslides: evidence from a large multi-temporal landslide inventory in Hong Kong. *Eng Geol* 292:106266. <https://doi.org/10.1016/j.enggeo.2021.106266>
- Ghasemi F, Neysiani BS, Nematbakhsh N (2020) Feature selection in pre-diagnosis heart coronary artery disease detection: a heuristic approach for feature selection based on information gain ratio and Gini index. In 2020 6th International Conference on Web Research. Tehran, Iran, pp 27–32. <https://doi.org/10.1109/ICWR49608.2020.9122285>
- Gong W, Juang CH, Wasowski J (2021) Geohazards and human settlements: Lessons learned from multiple relocation events in Badong, China—engineering geologist’s perspective. *Eng Geol* 285:106051. <https://doi.org/10.1016/j.enggeo.2021.106051>
- Gong W, Hu M, Zhang Y, Tang H, Liu D, Song Q (2022) GIS-based landslide susceptibility mapping using ensemble methods for Fengjie County in the Three Gorges Reservoir Region. *China Int J Environ Sci Technol* 19(8):7803–7820. <https://doi.org/10.1007/s13762-021-03572-z>
- Haque U, Da Silva PF, Devoli G, Pilz J, Zhao B, Khaloua A, Wilopo W, Andersen P, Lu P, Lee J, Yamamoto T, Keellings D, Wu J, Glass GE (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). *Sci Total Environ* 682:673–684. <https://doi.org/10.1016/j.scitotenv.2019.03.415>
- Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, New York
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *CATENA* 165:520–529. <https://doi.org/10.1016/j.catena.2018.03.003>
- Jenks GF (1967) The data model concept in statistical mapping. *International Yearbook of Cartography* 7:186–190
- Jordanova G, Gariano SL, Melillo M, Peruccacci S, Brunetti MT, Jemec Aulfič M (2020) Determination of empirical rainfall thresholds for shallow landslides in Slovenia using an automatic tool. *Water* 12(5):1449. <https://doi.org/10.3390/w12051449>
- Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11(3):425–439. <https://doi.org/10.1007/s10346-013-0391-7>
- Ko FWY, Lo FLC (2016) Rainfall-based landslide susceptibility analysis for natural terrain in Hong Kong—a direct stock-taking approach. *Eng Geol* 215:95–107. <https://doi.org/10.1016/j.enggeo.2016.11.001>
- Lam CLH, Lau JWC, Chan HW (2012) Factual report on Hong Kong rainfall and landslides in 2008. GEO Report No. 273. Geotechnical Engineering Office, Hong Kong Special Administration Region. [https://www.cedd.gov.hk/eng/publications/geo/geo-reports/geo\\_rpt273/index.html](https://www.cedd.gov.hk/eng/publications/geo/geo-reports/geo_rpt273/index.html)
- Lee S, Pradhan B (2007) Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4(1):33–41. <https://doi.org/10.1007/s10346-006-0047-y>
- Leung JK, Law TC (2002) Kriging analysis on Hong Kong rainfall data. *HKIE Transactions* 9(1):26–31. <https://doi.org/10.1080/1023697X.2002.10667865>
- Liu LL, Zhang YL, Xiao T, Yang C (2022) A frequency ratio-based sampling strategy for landslide susceptibility assessment. *Bull Eng Geol Environ* 81(9):360. <https://doi.org/10.1007/s10064-022-02836-3>
- Luti T, Segoni S, Catani F, Munafò M, Casagli N (2020) Integration of remotely sensed soil sealing data in landslide susceptibility mapping. *Remote Sens* 12(9):1486. <https://doi.org/10.3390/rs12091486>
- Maunsell-Fugro Joint Venture (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
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5(1):3–30. <https://doi.org/10.1002/hyp.3360050103>
- Ng CWW, Yang B, Liu ZQ, Kwan JSH, Chen L (2021) Spatiotemporal modelling of rainfall-induced landslides using machine learning. *Landslides* 18:2499–2514. <https://doi.org/10.1007/s10346-021-01662-0>
- O’Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41:673–690. <https://doi.org/10.1007/s11135-006-9018-6>
- Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *J Asian Earth Sci* 64:180–197. <https://doi.org/10.1016/j.jseaes.2012.12.014>
- Pradhan AMS, Lee SR, Kim YT (2019) A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan. *Korea Landslides* 16(3):647–659. <https://doi.org/10.1016/j.jseaes.2012.12.014>
- Rosi A, Peternel T, Jemec-Aulfič M, Komac M, Segoni S, Casagli N (2016) Rainfall thresholds for rainfall-induced landslides in Slovenia. *Landslides* 13(6):1571–1577. <https://doi.org/10.1007/s10346-016-0733-3>
- 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>
- Segoni S, Tofani V, Rosi A, Catani F, Casagli N (2018) Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Front Earth Sci* 85. <https://doi.org/10.3389/feart.2018.00085>
- Segoni S, Pappafico G, Luti T, Catani F (2020) Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization. *Landslides* 17:2443–2453. <https://doi.org/10.1007/s10346-019-01340-2>
- Sun D, Xu J, Wen H, Wang D (2021) Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: a comparison between logistic regression and random forest. *Eng Geol* 281:105972. <https://doi.org/10.1016/j.enggeo.2020.105972>
- Wang N, Lombardo L, Gariano SL, Cheng W, Liu C, Xiong J, Wang R (2021a) Using satellite rainfall products to assess the triggering conditions for hydro-morphological processes in different geomorphological settings in China. *Int J Appl Earth Obs Geoinf* 102:102350. <https://doi.org/10.1016/j.jag.2021.102350>
- Wang H, Zhang L, Luo H, He J, Cheung RWM (2021b) AI-powered landslide susceptibility assessment in Hong Kong. *Eng Geol* 288:106103. <https://doi.org/10.1016/j.enggeo.2021.106103>
- Wu X, Niu R, Ren F, Peng L (2013) Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges. *China Environ Earth Sci* 70(3):1307–1318. <https://doi.org/10.1007/s12665-013-2217-2>
- Xiao T, Zhang LM, Cheung RWM, Lacasse S (2022) Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques. *Geotechnique* 1–17. <https://doi.org/10.1680/jgeot.21.00160>
- Yao X, Tham LG, Dai FC (2008) Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong. *China Geomorphology* 101(4):572–582. <https://doi.org/10.1016/j.geomorph.2008.02.011>
- Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61(4):821–836. <https://doi.org/10.1007/s12665-009-0394-9>
- Zhang Y, Yan Q (2022) Landslide susceptibility prediction based on high-trust non-landslide point selection. *ISPRS Int J Geo-Inf* 11(7):398. <https://doi.org/10.3390/ijgi11070398>
- Zhang Y, Ayyub BM, Gong W, Tang H (2023) Risk assessment of roadway networks exposed to landslides in mountainous regions—a case study in Fengjie County, China. *Landslides* 1–13. <https://doi.org/10.1007/s10346-023-02045-3>
- Zhao F, Meng X, Zhang Y, Chen G, Su X, Yue D (2019) Landslide susceptibility mapping of karakorum highway combined with the application of SBAS-InSAR technology. *Sensors* 19(12):2685. <https://doi.org/10.3390/s19122685>



---

Zhou C, Yin K, Cao Y, Ahmed B, Li Y, Catani F, Pourghasemi HR (2018) Landslide susceptibility modelling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. *Comput Geosci* 112:23–37. <https://doi.org/10.1016/j.cageo.2017.11.019>

---

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

---

**Tianhe Ren · Liang Gao** (✉)

State Key Laboratory of Internet of Things for Smart City and Department of Civil and Environment Engineering, University of Macau, Macao 999078, China  
Email: gaoliang@um.edu.mo

**Tianhe Ren · Wenping Gong** (✉)

Faculty of Engineering, China University of Geosciences, Wuhan 430074, Hubei, China  
Email: wenpinggong@cug.edu.cn