

Landslide Susceptibility Mapping Using Feature Fusion-Based CPCNN-ML in Lantau Island, Hong Kong

Yangyang Chen^{1b}, Dongping Ming^{1b}, Xiao Ling, Xianwei Lv^{1b}, and Chenghu Zhou

Abstract—Landslide susceptibility mapping (LSM) is an effective way to predict spatial probability of landslide occurrence. Existing convolutional neural network (CNN)-based methods apply self-built CNN with simple structure, which failed to reach CNN's full potential on high-level feature extraction, meanwhile ignored the use of numerical predisposing factors. For the purpose of exploring feature fusion based CNN models with greater reliability in LSM, this study proposes an ensemble model based on channel-expanded pre-trained CNN and traditional machine learning model (CPCNN-ML). In CPCNN-ML, pre-trained CNN with mature structure is modified to excavate high-level features of multichannel predisposing factor layers. LSM result is generated by traditional machine learning (ML) model based on hybrid feature of high-level features and numerical predisposing factors. Lantau Island, Hong Kong is selected as study area; temporal landslide inventory is used for model training and evaluation. Experimental results show that CPCNN-ML has ability to predict landslide occurrence with high reliability, especially the CPCNN-ML based on random forest. Contrast experiments with self-built CNN and traditional ML models further embody the superiority of CPCNN-ML. It is worth noting that coastal regions are newly identified landslide-prone regions compared with previous research. This finding is of great reference value for Hong Kong authorities to formulate appropriate disaster prevention and mitigation policies.

Index Terms—Convolutional neural networks (CNNs), landslide susceptibility mapping (LSM), Lantau Island, remote sensing.

I. INTRODUCTION

AS ONE of the most destructive and frequently appeared geohazard, landslides have brought a tremendous threat

to the safety of human lives and possessions. From January 2004 to December 2016, more than 55 997 people were killed by landslides and landslide-induced geohazard [1]. Landslides are also responsible for the damage to infrastructure, including but not limited to buildings, roads, bridges, etc. The negative impact caused by landslides on the sustainable development of social and environment is cause for great concern. In order to mitigate landslide hazards, it is urgent to establish a system to predict where landslides would occur in the future [2]. Landslide susceptibility mapping (LSM) could give a spatial probability of landslide occurrence in a certain area [3], [4] and has been widely used in predicting landslide occurrence over the past decades. LSM provides an effective way to achieve "Reduce the Adverse Effects of Natural Disasters" target of sustainable development goal (SDG) 11. Based on LSM, landslide-prone areas could be identified, which helps the government to devise an effective disaster mitigation strategy [5] and accelerate the implementation of SDGs 2030.

Quantitative LSM methods include physical and statistical methods [3], [6]. Based on a slope stability study, physical methods predict landslide occurrence by modeling landslides' physical formation process [7] and can derive more accurate LSM results than statistical methods [8]. But, detailed geological field surveys are essential for precise physical modeling, which are extremely hard to obtain on a large scale. Therefore, physical methods are only suitable for the small area [9]; statistical methods weigh the contribution degree of landslide predisposing factors using statistical or regression models, based on which potential relationships between predisposing factors and landslide occurrence are excavated [10]. Statistical methods are data-driven and have become the mainstream for LSM. Landslide inventory and predisposing factors are the foundation of carrying out statistical methods. As geographic information system advances and easier acquisition of remote sensing products, landslide inventory and predisposing factors can be acquired efficiently on a large-scale, which is less time-consuming compared with field research [11] and, therefore, promote in-depth research in statistical methods. Existing statistical methods are divided into classical statistical and machine learning (ML)-based methods [12]. Frequency ratio [13], [14], weight of evidence [15] and information value [16], [17] are classical statistical models, which depend on linear relationships between landslide occurrence and predisposing factors. Considering the mechanism of landslide occurrence is complex, mining

Manuscript received August 10, 2020; revised January 27, 2021; accepted March 10, 2021. Date of publication March 17, 2021; date of current version April 9, 2021. This work was supported in part by the China Geological Survey under Grant DD20191006, in part by the National Natural Science Foundation of China under Grant 41671369, in part by the National Key Research and Development Program under Grant 2017YFB0503600-05, and in part by the Fundamental Research Funds for the Central Universities. (Corresponding author: Dongping Ming.)

Yangyang Chen, Dongping Ming, and Xiao Ling are with the School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China (e-mail: jimmyxiyangyang@hotmail.com; mingdp@cugb.edu.cn; lingx0527@163.com).

Xianwei Lv is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: xianweilv@whu.edu.cn).

Chenghu Zhou is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: zhouch@lreis.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3066378

nonlinear relationships between landslide inventory and predisposing factors are direction for the improvement of LSM reliability [18]; in recent years, ML models, such as logistic regression (LR) [19], [20], multilayer perceptron (MLP) [17], [21], [22], support vector machine (SVM) [23], [24], random forest (RF) [25]–[27], decision tree [28], [29], etc., have been widely used in LSM. Research has shown that ML-based methods are superior to classical statistical methods with higher prediction accuracy [30], for nonlinear relationships can be modeled effectively [31], [32]. To be noticed, the aforementioned ML models only utilize numerical information of predisposing factors, while ignoring feature extraction and spatial pattern representation. Hence, there is still much room for the improvement of LSM reliability.

With the rapid development of computer technology, deep learning (DL) models outperform traditional ML models in image classification [33], object detection [34], speech recognition [35], etc. DL models possess evident advantages in performance as high-level features of a large dataset can be fully excavated, based on which nonlinear relationships can be represented sufficiently. As the most mature DL framework, convolutional neural networks (CNNs) have been widely used in geoscience domain, such as scene classification [36], land-cover classification [37]–[40], lithological facies classification [41], [42], functional zone division [43], and ground target detection [44]–[46]. In recent years, CNN-based methods have been applied in the landslide-related domain, especially in landslide detection [47]–[56]. To the best of authors' knowledge, only a few research use CNN-based methods for LSM [57]–[59]. Wang *et al.* [57] first introduced CNN into LSM, achieves better LSM results than SVM and MLP. Fang *et al.* [58] utilize ML classifiers for LSM based on 1D-CNN-extracted features. Pham *et al.* [59] improve CNN-based LSM models by using a swarm intelligence algorithm for parameters optimization that yields an increase in LSM accuracy.

All the CNN-based LSM researches mentioned above have made great progress, but applied self-built CNN with simple architecture and complicated high-level features cannot be fully exploited in LSM. Thankfully, a pre-trained CNN model (PCNN) with a mature structure alleviates that problem in principle. Compared with self-built CNN, PCNN performs better in various domains, which can reach CNN's full potential on high-level feature extraction. Besides, using mature architecture and pre-trained parameters could weaken CNN's dependence on a large-scale training dataset. Moreover, applying PCNN can also save substantial effort on CNN training and hyper-parameters optimization. However, most PCNN can only process data with three channels and cannot be applied in LSM based on more than three types of predisposing factors.

Besides, existing CNN-based methods only utilize CNN-extracted high-level features while ignoring the usage of numerical predisposing factors. For traditional ML models, numerical predisposing factors have stood the test in LSM. Although numerical predisposing factor based traditional ML models are inferior to high-level feature-based CNN models, numerical predisposing factors still have unique potential and value in LSM. Given above, exploring a way to apply PCNN in LSM, in the

meantime, effectively fusing CNN-extracted high-level features with numerical landslide predisposing factors is a promising way to further improve the CNN-based LSM method.

Aiming at solving the above issues, this article proposed a feature fusion based ensemble model combining channel-expanded PCNN and traditional machine learning model (CPCNN-ML). CPCNN-ML and temporal landslide inventory were used for LSM in Lantau Island, Hong Kong (HK). Overall accuracy (OA), receiver operating characteristic (ROC) curve, area under the curve (AUC), and Wilcoxon signed-rank test (WSRT) were used to compare the performance of CPCNN-ML and benchmark models.

II. STUDY AREA

Lantau Island is the biggest outlying island of HK, China. As shown in Fig. 1, Lantau Island is located in the southwest of HK, covering an area of 147.16 km². It is a mountainous region with an elevation ranges from sea level to 934 m, where Lantau Peak is the highest point. As over 50% of the area consists of forest parks, Lantau Island has abundant and diverse indigenous forest resources. The Island has a relatively low population density compared with HK Island and Kowloon. There are about 1 05 000 citizens are settled on Lantau Island by 2011.

Reasons for selecting Lantau Island as a study area are as follows: First of all, Lantau Island enjoys monsoonal and sub-tropical weather with high average annual precipitation, seasonal rainfall triggers a considerable number of landslides every year; Second, even though Lantau Island has a low population density, there are several landmarks with a large stream of people, especially foreign tourists, which include HK International Airport, HK Disneyland Resort and Ngong Ping. Thus, the threat of landslides occurrence to lives and property in Lantau Island has to be taken seriously; third, for the purpose of monitoring and analyzing nature hazards, Geotechnical Engineering Office (GEO), Civil Engineering and Development Department (CEDD), Government of HK established a temporal Enhanced Natural Terrain Landslide Inventory (ENTLI) [60]. ENTLI is compiled by human interpretation of aerial photographs and constitutes almost all landslides in HK territory [61]. Therefore, ENTLI offers an opportunity for reliable LSM modeling and evaluation. The latest version of ENTLI contains 5340 landslides that occurred from 1974 to 2015 in Lantau Island, as shown in Fig. 2.

III. METHODOLOGY AND EXPERIMENTS

The methodology flowchart of LSM using CPCNN-ML takes five major steps (see Fig. 3). First, various predisposing factors are collected and preprocessed to predisposing factor layers. Then, the dataset for model training, validation, performance evaluation, and LSM are generated based on ENTLI (1994–2015). Next, by adjusting the architecture and fine-tune strategy, PCNN was modified as a channel-expanded PCNN model (CPCNN) to extract high-level features of predisposing factor layers. After that, CPCNN-extracted high-level features are fused with numerical predisposing factors into hybrid features

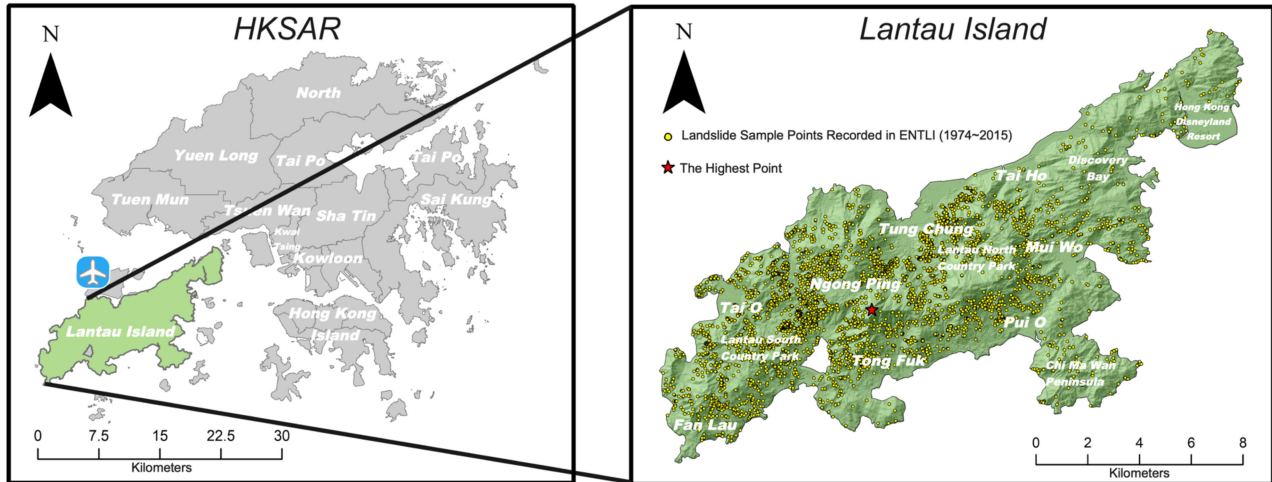


Fig. 1. Location of Lantau Island.

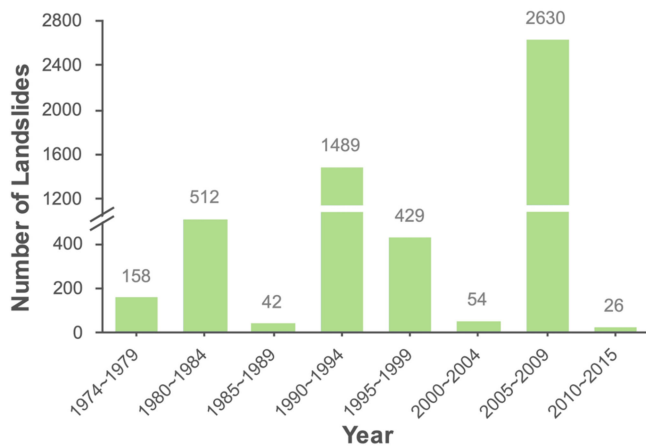


Fig. 2. The number of landslides occurred in Lantau Island.

and fed into the traditional ML model for LSM. Finally, the performance of CPCNN-ML is evaluated using multiple statistical evaluation methods.

A. Data Acquisition and Preprocessing

1) *Landslide Predisposing Factors*: Based on previous LSM experience in Lantau Island [62], 10 landslide predisposing factors were utilized in this study (see Fig. 4), including elevation, slope, aspect, profile curvature, plan curvature, distance to fault, lithology, topographic wetness index (TWI), normalized difference vegetation index (NDVI), and land cover. Predisposing factors were processed via ArcGIS 10.3 and classified into four categories: morphological, geological, hydrological, and land cover-related.

Morphological predisposing factors are of significant effective in LSM. Elevation, slope, aspect, profile curvature, and plan curvature are five of the most commonly used morphological predisposing factors, which were calculated using the digital elevation model (DEM). DEM (30 m per-pixel) was provided

by Esri China (HK); as geological predisposing factors, lithological, and fault lines were extracted from a simplified geological map, which was also derived from Esri China (HK). Distance to fault was obtained by calculating Euclidean Distance from fault; in Lantau Island, ground water flow and soil moisture are dominant hydrological predisposing factors, which were represented by TWI and can be calculated from DEM using hydrology tools in ArcGIS; land cover is indispensable in LSM research, for human activity along steep mountainous region may destabilize natural terrain [63]. For this consideration, the land cover was taken as related landslide predisposing factors. Land cover was obtained from FROM-GLC product (30 m per-pixel) [64], which has seven Level I land cover types. NDVI, as an indicator to measure vegetation coverage rate, is another important predisposing factor used for representing land cover situation. NDVI was calculated using Landsat TM image with 30-m spatial resolution, which was taken on March 14, 2006.

Raw predisposing factors cannot be used directly no matter which LSM model is applied, as different predisposing factors have different dimensions as well as ranges and types. Therefore, normalization and reclassification of predisposing factors are indispensable. As shown in Table I, predisposing factors were reclassified by reference to the experience of Huang *et al.* [62]. After that, predisposing factors were normalized in the range from 0 to 255 using min-max normalization. Finally, all reclassified predisposing factors were composed of predisposing factor layers and resampled to 30 m per-pixel.

2) *Establishment of Dataset*: Location of each landslide recorded in ENTLI is represented by a point shapefile. Considering that early landslide records may not be reliable and have a low correlative degree with recently occurred landslides, only 3196 landslides occurred from 1994 to 2015 were used for dataset establishment. Generation of nonlandslides sample points is equally important in order to implement supervised learning algorithms. In this study, 3196 nonlandslide sample points were randomly generated in accordance with three criteria: Nonlandslide sample points should be at least 50 m away from the border; nonlandslide sample points should be 200 m

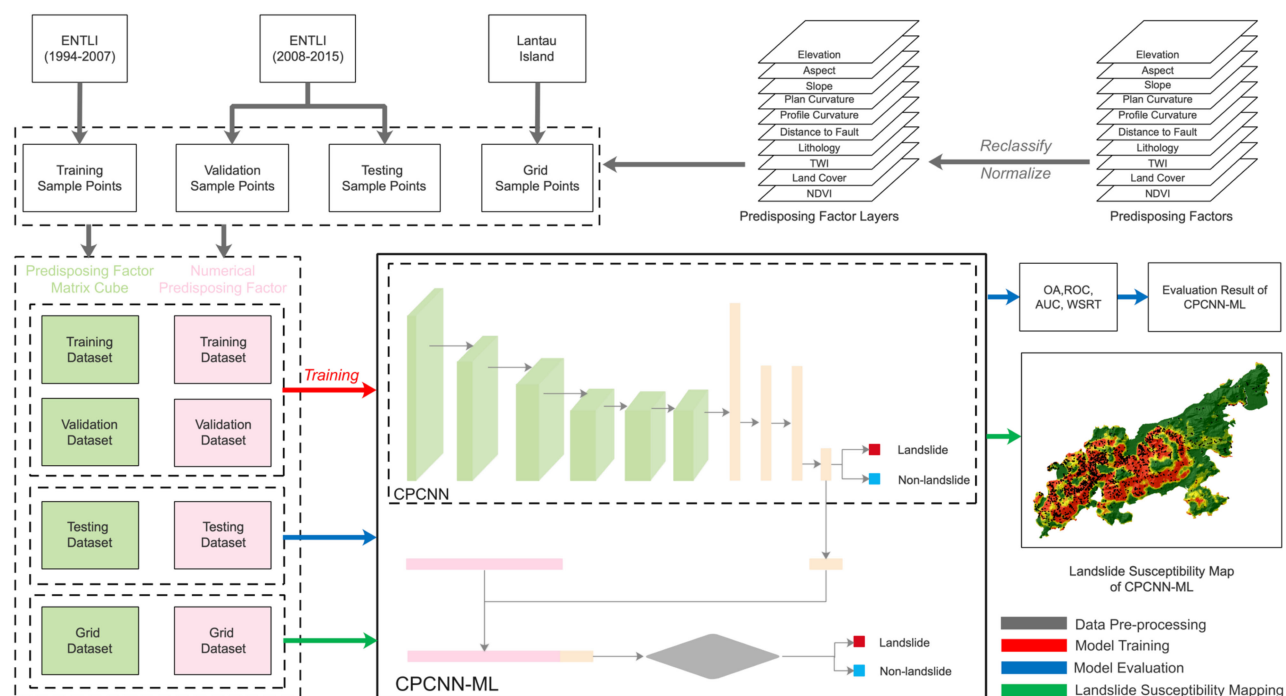


Fig. 3. Flowchart of methodology.

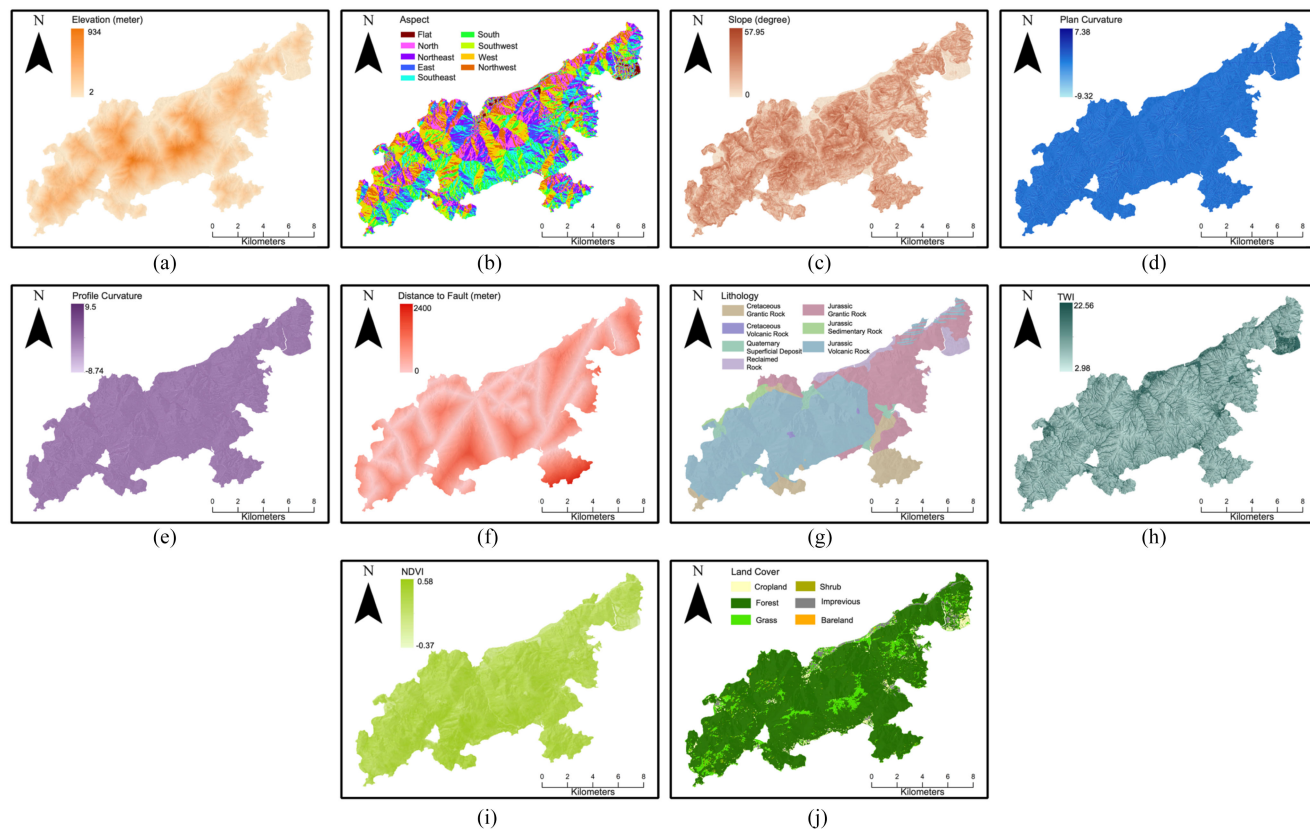


Fig. 4. Landslide predisposing factors. (a) Elevation. (b) Aspect. (c) Slope. (d) Plan curvature. (e) Profile curvature. (f) Distance to fault. (g) Lithology. (h) TWI. (i) NDVI. (j) Land cover.

TABLE I
DESCRIPTION OF RECLASSIFIED PREDISPOSING FACTORS

Predisposing Factors	Class	Category/Range of Value	Grey Value of Predisposing Factor Layer
Morphological	Elevation	1 2m~100m	28
		2 100m~200m	56
		3 200m~300m	84
		4 300m~400m	112
		5 400m~500m	140
		6 500m~600m	168
		7 700m~800m	196
		8 800m~900m	224
		9 900m~934m	252
	Aspect	1 Flat	28
		2 North	56
		3 Northeast	84
		4 East	112
		5 Southeast	140
		6 South	168
		7 Southwest	196
		8 West	224
		9 Northwest	252
	Slope	1 0°~2°	36
		2 2°~5°	72
		3 5°~15°	108
		4 15°~25°	144
		5 25°~35°	180
		6 35°~55°	216
		7 55°~57.95°	252
	Plan Curvature	1 -9.32~-8	28
		2 -8~-6	56
		3 -6~-4	84
		4 -4~-2	112
		5 -2~0	140
		6 0~2	168
		7 2~4	196
		8 4~6	224
		9 6~7.38	252
	Profile Curvature	1 -8.74~-8	25
		2 -8~-6	50
		3 -6~-4	75
		4 -4~-2	100
		5 -2~0	125
		6 0~2	150
		7 2~4	175
		8 4~6	200
		9 6~8	225
		10 8~9.5	250
Geological	Distant to Fault	1 0~400m	42
		2 400~800m	84
		3 800~1200m	126
		4 1200~1600m	168
		5 1600~2000m	210
		6 2000~2400m	252
	Lithology	1 Reclaimed Rock	36
		2 Quaternary Superficial Deposit	72
		3 Cretaceous Volcanic Rock	108
		4 Cretaceous Granitic Rock	144
		5 Jurassic Volcanic Rock	180
		6 Jurassic Sedimentary Rock	216
		7 Jurassic Granitic Rock	252
	Hydrological	1 2.98~6	36
		2 6~9	72
		3 9~12	108
		4 12~15	144
		5 15~18	180
		6 18~21	216
		7 21~22.56	252
Land-cover Related	Land Cover	1 Cropland	42
		2 Forest	84
		3 Grass	126
		4 Shrub	168
		5 Imperivous	210
		6 Bareland	252
	NDVI	1 -0.37~-0.3	25
		2 -0.3~-0.2	50
		3 -0.2~-0.1	75
		4 -0.1~0	100
		5 0~0.1	125
		6 0.1~0.2	150
		7 0.2~0.3	175
		8 0.3~0.4	200
		9 0.4~0.5	225
		10 0.5~0.58	250

TABLE II
DESCRIPTION OF DATASET

Type	Source (Period)	Amount of Landslide Sample Points	Amount of Non-landslide Sample Points
Training Data (70%)	ENTLI (1994~2007)	656	656
Validation Data (30%)	ENTLI (2008~2015)	281	281
Test Data		2259	2259

away from the nearest landslide sample point; distance between any two nonlandslide sample points should be at least 50 m.

Landslides inventory was divided into training, validation, and testing sample points based on temporal sequence, which is more reasonable than stochastic partition adopted by most LSM research. As shown in Table II and Fig. 5, 656 landslides occurred from 1994 to 2007 composed training sample points with 656 randomly selected nonlandslide sample points, which were used for model training. Landslides occurred from 2008 to 2015 were used for model validation and performance evaluation. Validation sample points contain randomly selected 281 landslide sample points and 281 nonlandslide sample points. The other 2259 landslide sample points and 2259 nonlandslide sample points made up testing sample points for accuracy evaluation of the final LSM result.

Applicable data forms for CNN and traditional ML models are predisposing factor matrix cube and numerical predisposing factors, respectively. The process of generating predisposing factor matrix cube and numerical predisposing factors is given in Fig. 6. Predisposing factor layers needed to be clipped into $S \times S \times 10$ predisposing factor matrix cubes by sample points,

where S stands for the scale parameter of CNN dataset. Each clipped $S \times S \times 10$ matrix cube shares the same geometric center with a corresponding sample point. Predisposing matrix cubes and labels of corresponding sample points constituted a dataset for CNN models. Grey value of sample points on predisposing factor layers and samples' label constituted dataset for traditional ML models.

B. Convolutional Neural Network

As the representative of DL, CNN can be described as a certain kind of MLP with a convolutional (Conv) and pooling structure. CNN is constituted of multiple Conv layers, pooling layers, fully-connected (FC) layers, and classifier. Conv layers are used for high-level feature extraction, which is the core of CNN. With the help of parameter sharing and sparse local connectivity scheme, input data x is abstracted to a high-level feature map h after passing through the Conv layer. Weights of Conv kernel W and corresponding bias term b are optimized during the training process. The k th Conv layer is defined as [65]

$$g_n^k = f \left(\sum_m W_{mn}^k \otimes x_m^k + b_n^k \right) \quad (1)$$

where g_n^k is the n th high-level feature map extracted by the k th Conv layer. $f(\cdot)$ stands for the nonlinear activation function and $\otimes$ is Conv operation. x_m^k is the m th layer of input data. W_{mn}^k denotes the weights of the n th Conv kernel with x_m^k and b_n^k is the bias term.

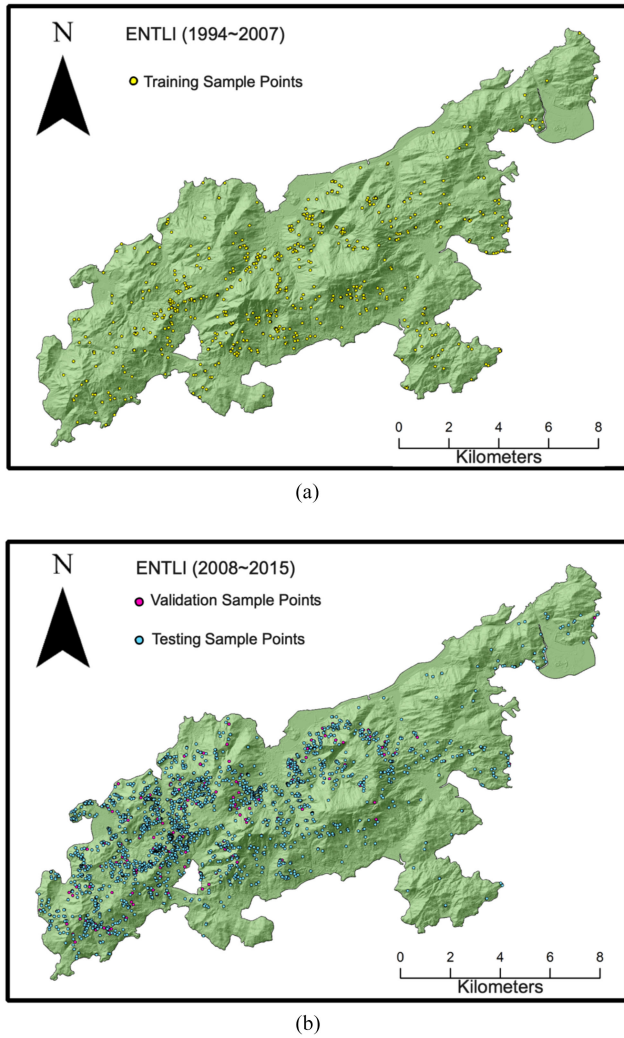


Fig. 5. Distribution of landslide sample points. (a) Training sample points. (b) Validation and testing sample points.

Pooling layers reduce the parameter number of the following network layers. Applying pooling operation after Conv layer can not only raise training efficiency but can also avoid over-fitting. Max pooling is one of the commonly used down-sampling operation, which utilizes the maximum value of quadratic regions of feature map as [66]

$$h_n^k = \Phi(g_n^k) \quad (2)$$

where $\Phi(\cdot)$ stands for max-pooling operation. h_n^k is the n th feature map down-sampled by k th max-pooling layer.

The last pooling layer is followed by multiple FC layer. FC layer in CNN and MLP have the same mechanism, the j th FC layer is defined as [65]

$$j_k = f(W_k \times y_k) + b_k \quad (3)$$

where y_k stands for input data and j_k is output features of the k th FC layer. W_k and b_k stand for weights and biases, respectively. $f(\cdot)$ indicates activate function. To be noticed, for the first FC layer, input data are flattened 1-D feature map of the last pooling

layer. And output features of the last FC layer are fed into a classifier.

C. CPCNN-ML

With the aim of activating the performance of CNN with high efficiency, PCNN was first modified as CPCNN for high-level feature extraction of predisposing factor layers. As the handiest and most widely used PCNN model in various domains, pretrained AlexNet was selected as the basis of CPCNN model in this study. AlexNet [67] won the championship of LSVRC 2012 and, therefore, has remarkable ability of high-level feature extraction and representation. Original AlexNet consists of five Conv layers, three pooling layers, and three FC layers. The size of input data is limited to $227 \times 227 \times 3$ and filtered by 96 Conv kernels ($11 \times 11 \times 3$) in the 1st Conv layer. Input data of the 2nd Conv layer is filtered with 256 Conv kernels ($5 \times 5 \times 96$). For the 3rd and 4th Conv layer, input data are filtered by 384 Conv kernels ($3 \times 3 \times 256$) and 384 Conv kernels ($3 \times 3 \times 384$), respectively. The 5th Conv layer has 256 Conv kernels ($3 \times 3 \times 384$), and its output feature map is flattened into a 1-D feature. Both of the following two FC layers have 4096 neurons, whereas neuron number of the last FC layer is equal to the number of labels. Finally, features generated by the last FC layer are feed into softmax classifier for final classification result. AlexNet has succeeded by virtue of the innovative use of rectified linear units (ReLU) [68]

$$n(x) = \max(0, x) \quad (4)$$

where x is the input signal of ReLU. Using ReLU as an activation function can avoid vanishing gradient problems in training process and alleviates the problem of overfitting. Applying dropout regularization in the FC layer also avoids overfitting in the training process, which contributes to the improvement of classification accuracy.

To be noticed, most PCNN are initially designed for RGB image classification, including AlexNet. In other words, the original AlexNet and corresponding fine-tune strategy can only be used to process matrix cube with three channels. However, it is impossible to generate reliable LSM results with only three types of triggering factors. Therefore, by adjusting the structure, PCNN was modified as CPCNN to process predisposing factor matrix cube with more than three channels. As can be seen in Fig. 6, the 1st Conv layer of AlexNet was redesigned, size of 96 Conv kernels was modified from $11 \times 11 \times 3$ to $11 \times 11 \times 10$. Besides, the neuron number of the last FC layers was set as 2 in order to generate landslide and nonlandslide probability.

Considering that the size of the training dataset is too small to fully train CPCNN, fine-tune strategy is applied in the training process. Fine-tune is a retraining strategy for CNN which was pretrained by a large dataset. Applying fine-tune strategy could avoid the pitfalls of using a small dataset and improve the performance and efficiency of CNN. For the original AlexNet which was pretrained by ImageNet dataset, pre-trained parameters of all Conv layers were reloaded while parameters of FC layers were set as default during fine-tune process. For CPCNN, the

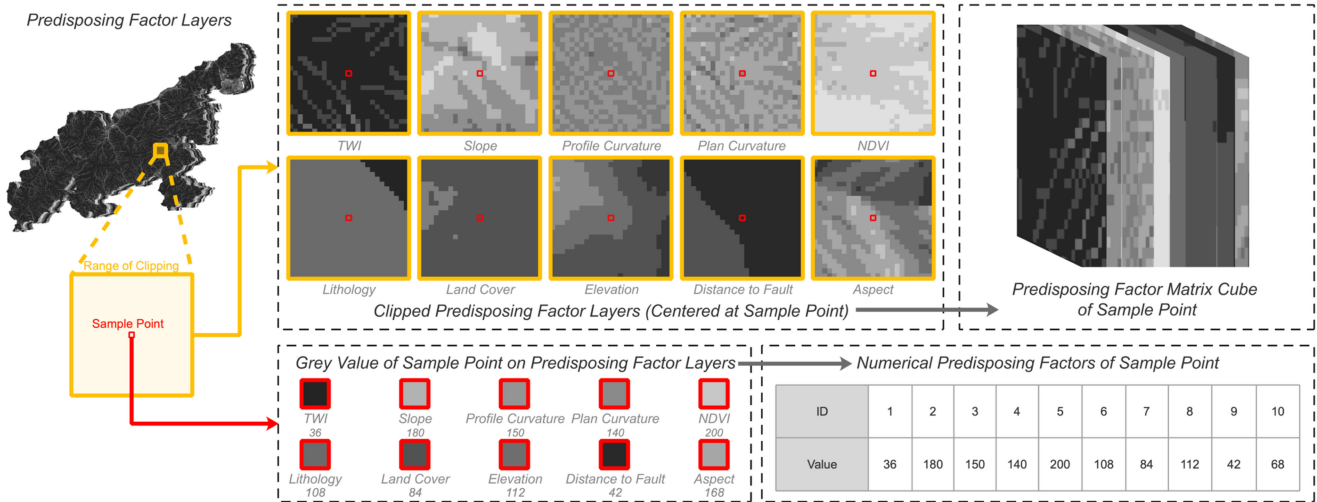


Fig. 6. Process of generating predisposing factor matrix cube and numerical predisposing factors.

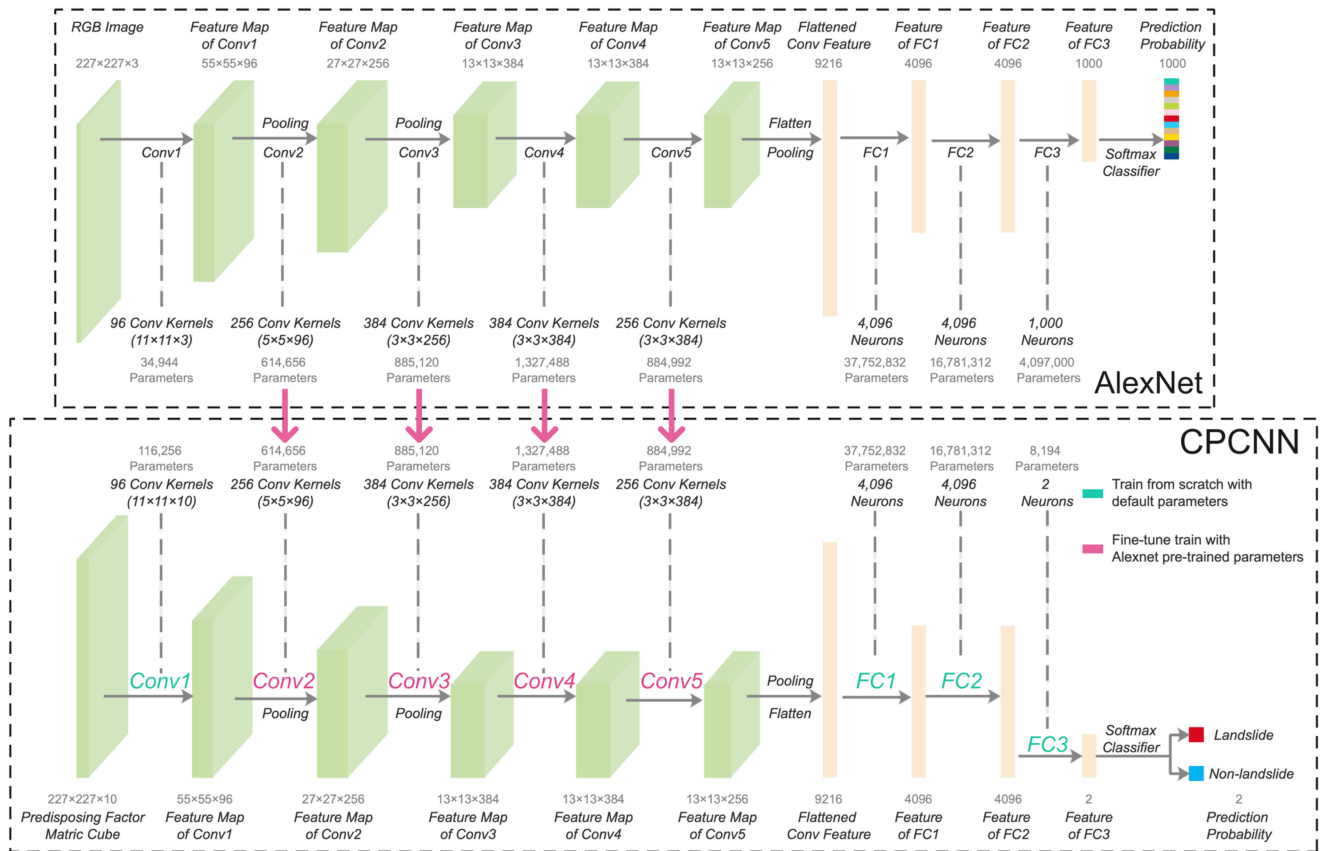


Fig. 7. Architecture and fine-tune strategy of CPCNN.

situation is different due to the adjustment of the AlexNet structure. As shown in Fig. 7, the parameter number of the 1st Conv layer is not matched with pretrained parameter number of the corresponding layer. Therefore, fine-tune strategy was modified in CPCNN; parameters of the 1st Conv layer were also set as default along with three FC layers. By doing so, the problem that PCNN cannot be applied in LSM is solved. Additionally, training the 1st Conv layer from scratch will helps CNN better adapt

to multichannel matrix cube in high-level feature extraction. After reloading the parameters of the 2nd, 3rd, 4th, and 5th Conv layers, landslide and nonlandslide datasets were resampled to $227 \times 227 \times 10$ and fed into CPCNN for network retraining. To adaptively learn high-level feature extraction rules suitable for predisposing factor matrix cube and LSM, parameters of all Conv layers and FC layers were trainable during the retraining process. After CNN training is completed, the Softmax classifier

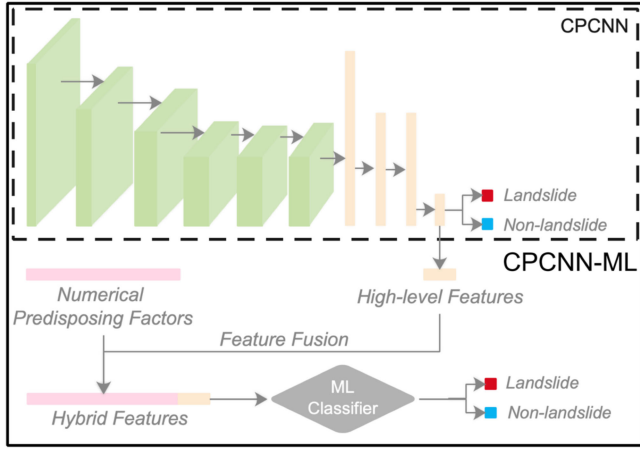


Fig. 8. CPCNN-ML for LSM.

cloud outputs the probability of landslide occurrence. The output of the 3rd FC layer contains high-level features extracted from predisposing factor layers.

In CPCNN-ML (see Fig. 8), extracted patch-based high-level features and point-based numerical landslide predisposing factors were fused into a 1-D feature vector, which is hybrid features of landslide predisposing factors. Finally, hybrid features were fed into the traditional ML model to predict the spatial probability of landslides occurrence.

D. LSM Using CPCNN-ML

Experiments of LSM were carried out on Windows 10 OS with 3.6 GHz Core i7-7700 and NVIDIA GeForce GTX 1080. Tensorflow-GPU 1.7.0 and Scikit-learn 0.22 were selected as DL and ML platforms, respectively.

Scale parameter S of the CNN dataset was set as 25. During the CNN training process, pre-trained parameters of AlexNet were first reloaded into CPCNN using aforementioned strategy. Next, CPCNN was retrained by training dataset using stochastic gradient descent, with a learning rate of 0.01 and dropout rate of 0.5. At the end of each training epoch, a validation dataset was used to provide an unbiased evaluation of model fit. With the help of validation loss, an early stopping strategy was applied to avoid overfitting. When the validation loss was no longer decrease, training was halted and parameters of the corresponding epoch were stored in the checkpoint file. Then, the training dataset was fed into CPCNN with stored parameters to extract high-level features of predisposing factors. Finally, CPCNN-extracted high-level features were fused with numerical predisposing factors, which were used for the training of traditional ML models.

For traditional ML models, RF, MLP, SVM, and LR were constructed, which have been used extensively in previous LSM research. Automated grid-search strategy is applied to select the appropriate hyper-parameters of traditional ML models. Number of trees, max depth, and max features of RF were set as 350, 7, and “sqrt,” respectively. For MLP, “sgd” solver was applied, the number of hidden layers was set as 200, initial learning was set as 0.005, and optimized adaptively during the training process. SVM was utilized based on “rbf” kernel, with

C of 0.9 and γ of 0.015. For LR, “lbfgs” solver and $L1$ penalty were selected.

After the training procedure ended, CPCNN-ML could map the landslide susceptibility of Lantau Island. In this experiment, Lantau Island was divided into 1 63 052 grid cells with 30-m spatial resolution. Landslide susceptibility of grid cells was classified into five grades using Jenks natural breaks algorithms: very high, high, moderate, low, and very low. Fig. 9 shows LSM results generated by CPCNN-ML.

The distribution of landslide susceptibility maps generated by CPCNN-ML based on RF (CPCNN-RF), CPCNN-ML based on MLP (CPCNN-MLP), CPCNN-ML based on SVM (CPCNN-SVM), and CPCNN-ML based on LR (CPCNN-LR) are roughly the same. Landslide-prone areas are mainly distributed in south-east Lantau Island while low landslide susceptibility areas are distributed in the northwest.

E. Performance Evaluation

To evaluate the predictive performance of CPCNN-ML, several statistical indicators including OA, ROC curve, and AUC were performed in SPSS Statistics 26. OA measures the percentage of correctly predicted testing sample points over all testing sample points. ROC curve is an effective and easy-to-use indicator that evaluates the performance of binary classifier and forecasting model. ROC curve plots false positive rate (FPR) on X-axis against true positive rate (TPR) on Y-axis. FPR and TPR can be calculated as [69]

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$TPR = \text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (7)$$

$$FPR = 1 - \text{Specificity} = \frac{FP}{FP + TN} \quad (8)$$

where true positive (TP) and true negative (TN) indicate the number of correctly predicted landslide and nonlandslide sample points, respectively. False positive (FP) and false negative (FN) are the numbers of mispredicted landslide and nonlandslide sample points. AUC stands for the area under the ROC curve and indicates the predictive performance of the model quantitatively, which ranges from 0.5 to 1. The higher the AUC value, the better the predictive performance of LSM models.

Testing and training dataset was processed by CPCNN-ML and probability of landslide occurrence was assigned to corresponding sample points. Evaluation results based on the training dataset are illustrated in Table III. Evaluation results based on the testing dataset are shown in Table IV and Fig 11.

As shown in Table III, CPCNN-ML obtained fairly good accuracy, especially the CPCNN-RF and CPCNN-MLP, which illustrates that the proposed CPCNN-ML fit well on the training dataset. Fig. 9 and Table IV show that CPCNN-ML achieved satisfactory LSM prediction results with OA value greater than 88% and AUC value greater than 0.94. Of the four CPCNN-ML

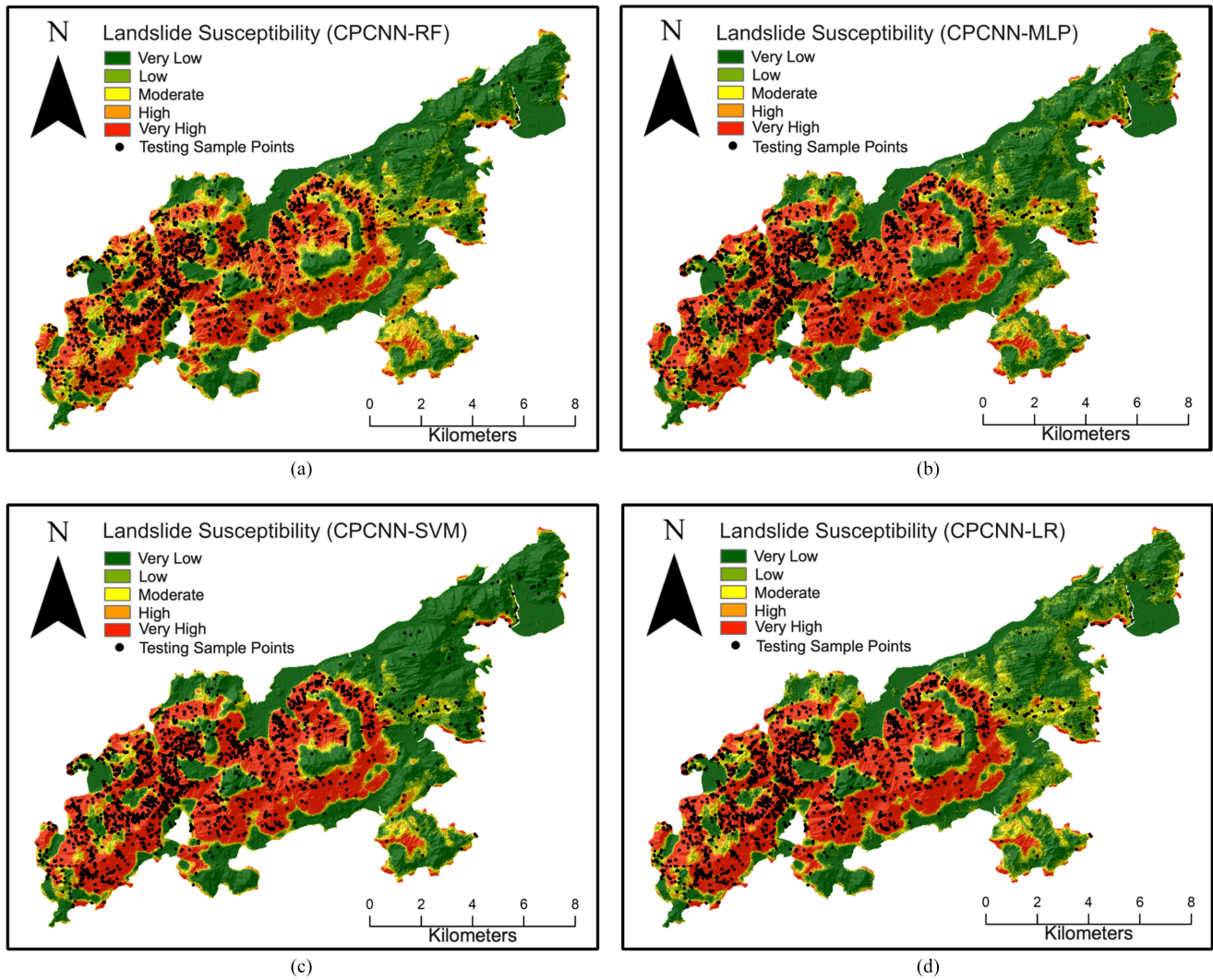


Fig. 9. LSM results of CPCNN-ML. (a) CPCNN-RF. (b) CPCNN-MLP. (c) CPCNN-SVM. (d) CPCNN-LR.

TABLE III
EVALUATION RESULT OF LSM MODELS BASED ON TRAINING DATASET

LSM Method		OA	AUC
CPCNN-ML	CPCNN-RF	93.75	0.988
	CPCNN-MLP	91.08	0.975
	CPCNN-SVM	89.25	0.963
	CPCNN-LR	89.86	0.961
CPCNN		88.41	0.962
3D-CNN [57]		85.06	0.921
ML	RF	88.57	0.949
	MLP	84.52	0.915
	SVM	86.89	0.934
	LR	73.78	0.815

TABLE IV
EVALUATION RESULT OF LSM MODELS BASED ON TESTING DATASET

LSM Method		OA	AUC
CPCNN-ML	CPCNN-RF	88.78	0.95
	CPCNN-MLP	88.29	0.947
	CPCNN-SVM	88.14	0.949
	CPCNN-LR	87.92	0.943
CPCNN		88.11	0.946
3D-CNN [57]		81.67	0.895
ML	RF	85.39	0.912
	MLP	82.78	0.896
	SVM	82.22	0.894
	LR	73.17	0.798

models, CPCNN-RF achieved the best landslide occurrence prediction ability with the highest OA value (88.78%) and AUC value (0.95). The predictive performance of CPCNN-MLP and CPCNN-SVM are comparable, which are slightly worse than CPCNN-RF. While CPCNN-LR has the worst predictive performance.

F. Model Comparison

Proposed CPCNN-ML was compared with various benchmark LSM models, including CPCNN, self-built CNN, RF, MLP, SVM, and LR. The structure of self-built CNN is referenced to 3D-CNN designed by Wang *et al.* [57]. LSM results of benchmark models are presented in Fig. 10. Table III shows the evaluation results of benchmark models based on the training

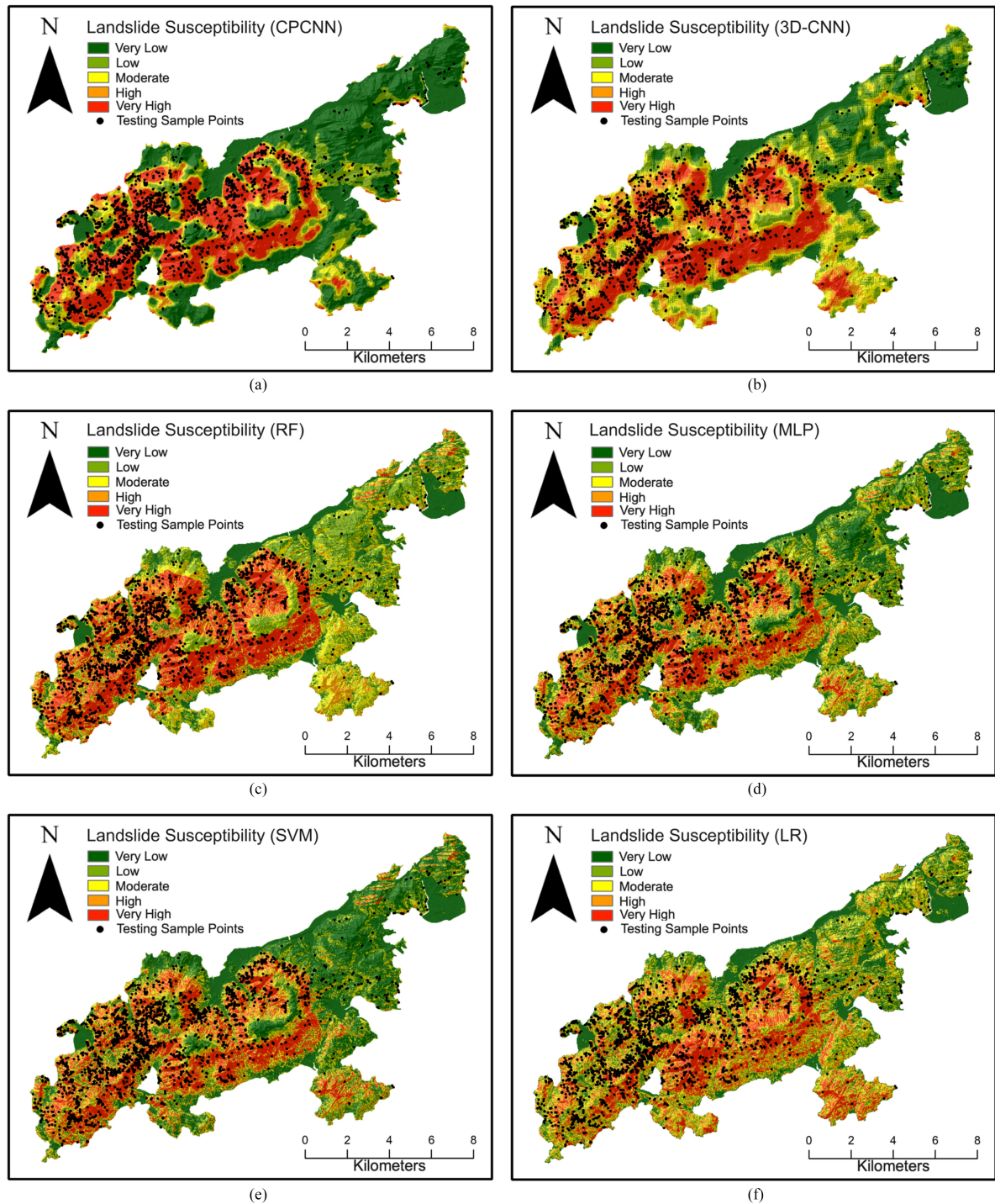


Fig. 10. LSM results of benchmark models. (a) CPCNN. (b) 3D-CNN [57]. (c) RF. (d) MLP. (e) SVM. (f) LR.

dataset. Fig. 11 and Table IV summarized evaluation results of benchmark models based on the testing dataset.

It can be seen in Figs. 9 and 10 that spatial distribution of landslide occurrence between CPCNN-ML and traditional ML models have differences. For RF, MLP, SVM, and LR, low and very low landslide susceptibility areas only occupied relatively small proportion of Lantau Island. The spatial distribution of CPCNN is much the same as CPCNN-ML, while areas labeled as high and very high susceptibility are larger.

Landslide susceptibility predictive ability of different models can be quantitatively compared based on Fig. 11 and Table IV. For traditional ML models, LR performed the worst with the lowest OA (73.78%) and AUC (0.815) accuracy, whereas RF and SVM performed much better. 3D-CNN achieved higher accuracy than MLP and LR, however, it showed no advantage over RF and SVM, and there is a certain gap between 3D-CNN and CPCNN. CPCNN-ML and CPCNN outperformed 3D-CNN and traditional ML models in terms of OA and AUC

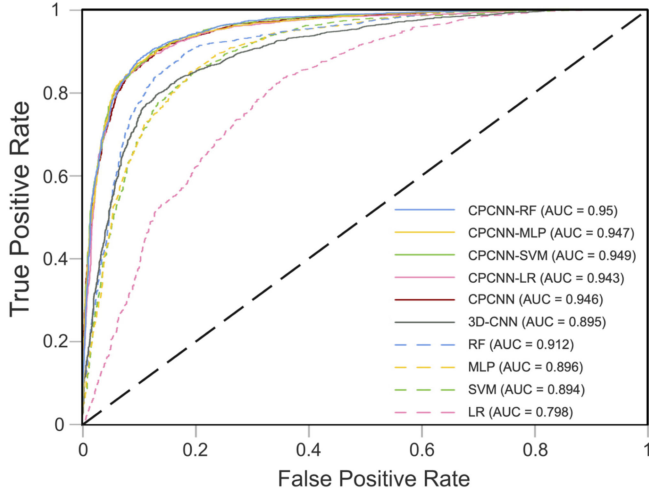


Fig. 11. ROC curve of different LSM models based on testing dataset.

TABLE V
WRST RESULTS AMONG CPCNN AND CPCNN-ML BASED ON TESTING DATASET

Pair-wise Comparison	Z Value	p Value	Significance
CPCNN-RF vs. CPCNN	-27.97	< 0.0001	Yes
CPCNN-MLP vs. CPCNN	-33.651	< 0.0001	Yes
CPCNN-SVM vs. CPCNN	-48.917	< 0.0001	Yes
CPCNN-LR vs. CPCNN	-38.983	< 0.0001	Yes

accuracy. Performance of CPCNN is behind CPCNN-RF, CPCNN-MLP, and CPCNN-SVM, with AUC of 0.946 and OA of 88.11%. Whereas CPCNN-LR has relatively poor predictive performance, which is even worse than CPCNN.

WSRT is a widely used nonparametric statistical hypothesis test for matched samples comparison, which is used to determine whether there is a statistical significant difference among the results of CPCNN-ML and CPCNN. WSRT was performed in SPSS Statistics 26, and Table V shows WSRT results based on the proposed CPCNN and CPCNN-ML. As shown in Table V, p value and Z value are less than 0.0001 and -1.96 , respectively, for all pair-wise comparisons. This means that null-hypothesis was rejected; the difference among LSM results of CPCNN and CPCNN-ML is significant.

IV. DISCUSSION

A. Effectiveness of CPCNN-Extracted High-Level features

With the aim of boosting performance and efficiency of self-built CNN, PCNN and corresponding fine-tune strategy were modified and applied in CPCNN-ML to extract high-level features of landslide predisposing factor layers. As shown in Fig. 11 and Table IV, CPCNN achieved significantly higher accuracy compared with traditional ML models, which proves the superiority of CPCNN-extracted high-level features. Thanks to fine-tune strategy and the application of mature PCNN architecture, the potential of CNN is fully aroused. By making full use of high-level features extracted from predisposing factor matrix cube, CPCNN is well ahead of self-built CNN in mapping landslide susceptibility. CPCNN achieved an AUC value of

0.946 and an OA value of 88.11% in Lantau Island, which is 0.051% and 6.44% higher than self-built 3D-CNN. Besides, effort spending on hyper-parameter optimization is reduced significantly using CPCNN. For self-built CNN, a large number of hyper-parameters needed to be optimized to gain satisfactory LSM results, including network structure parameters, scale of dataset, activation function, loss function, and optimizer. In contrast, the training process of CPCNN is more efficient, only one parameter need to be optimized, which is the size of predisposing factor matrix cube.

B. Effectiveness of Feature Fusion and Hybrid Model Strategy of CPCNN-ML

Traditional ML models and CPCNN have their own advantages and limitations. Landslide susceptibility map generated by CPCNN is good at integrity with high precision. Prediction accuracy of traditional ML models is lower than CPCNN, whereas landslide susceptibility maps generated by traditional ML models show more details.

Traditional ML models and CPCNN are based on numerical predisposing factors and high-level features, respectively. There are two main differences between these two features. First, the feature extraction process of CNN is a black box operation, which leads to the extracted high-level features without explicit physical or geoscience meaning. In contrast, the physical or geoscience meaning of numerical predisposing factors is well defined. Second, high-level features are patched-based and mainly focus on neighboring information of sample points, while numerical predisposing factors are point-based and mainly focuses on the sample point itself.

By adopting feature fusion operation, the traditional ML model and CPCNN are tightly integrated into CPCNN-ML. Both point-based numerical predisposing factors and patch-based high-level features can be used for LSM. Based on which, traditional ML models could fill in the details for LSM results generated by CPCNN. Besides, LSM results generated by CPCNN-ML could be explainable and more reliable.

From Fig. 11 and Table IV, it can be found that CPCNN-ML effectively combined the advantages of traditional ML models and CPCNN. Quantitatively, CPCNN-RF, CPCNN-MLP, and CPCNN-SVM are superior to both CPCNN and traditional ML models on AUC and OA accuracy. Especially the CPCNN-RF, 0.004 higher AUC and 0.68% higher OA are achieved compared with CPCNN. Qualitatively, the LSM result of CPCNN-ML is fine-grained, contains much more details than CPCNN. WSRT results also demonstrate that landslide susceptibility maps generated by CPCNN and CPCNN-ML are significantly different. Above all, CPCNN-ML has a better ability in generating reliable fine-grained LSM results with the help of feature fusion and hybrid model strategy.

C. Sensitive Analysis

As mentioned in Section IV-A, the size of predisposing factor matrix cube S is the only one hyper-parameter during the CNN training process of CPCNN-ML. Whether the selection of S is desirable or not will directly affect the quality of high-level

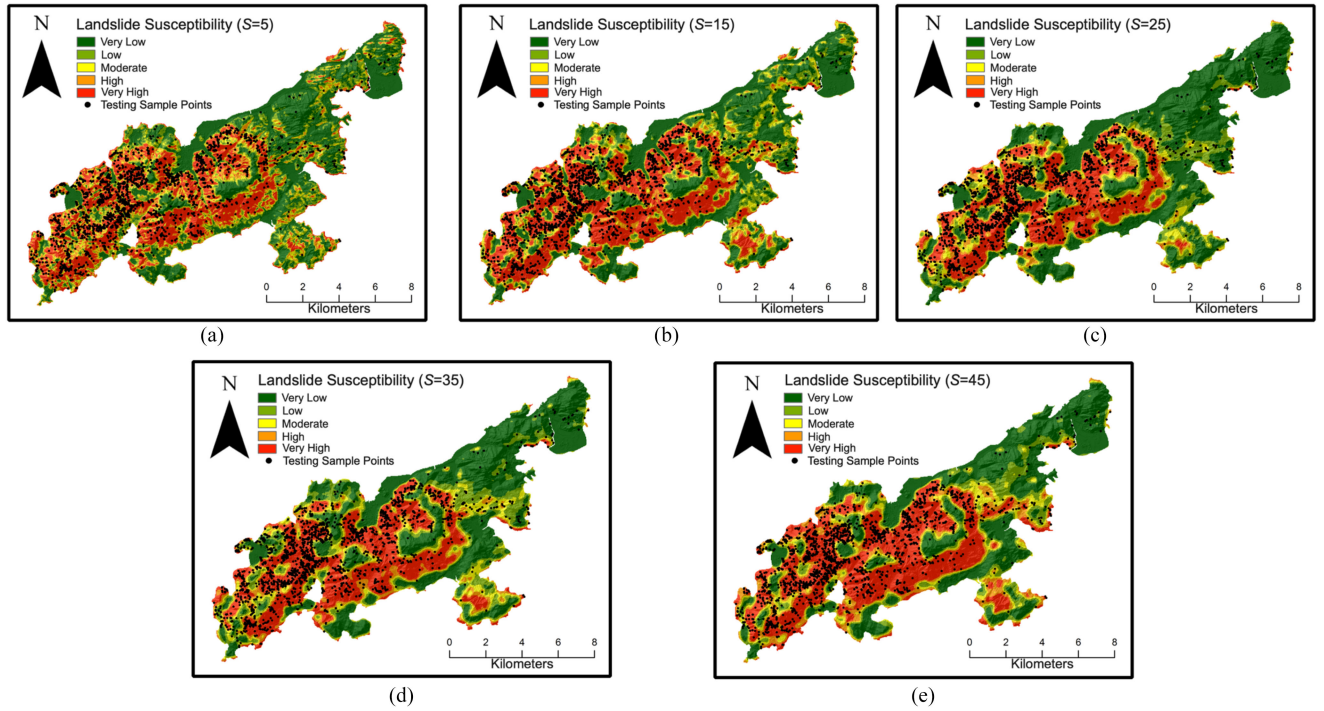


Fig. 12. LSM results generated by CPCNN with different S . (a) $S = 5$. (b) $S = 15$. (c) $S = 25$. (d) $S = 35$. (e) $S = 45$.

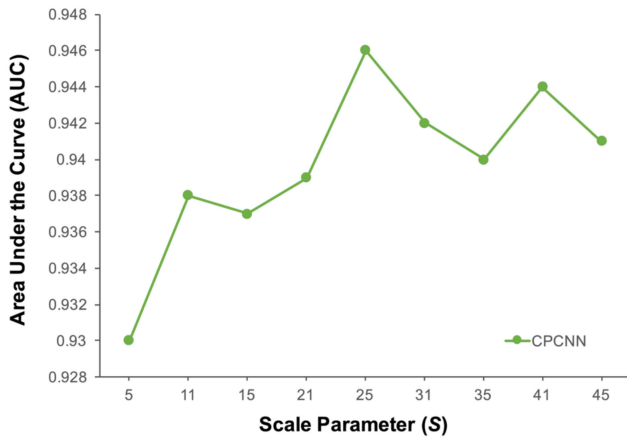


Fig. 13. AUC variation of CPCNN with different parameter S .

feature extraction, and thus influence the final LSM result. Considering that the CNN training process of CPCNN-ML and CPCNN is exactly the same, hyper-parameter's sensitivity of CPCNN was discussed in this section. Parameters S range from 5 to 45 were tested for comparison. Fig. 12 shows LSM results generated by CPCNN based on different S . Fig. 13 illustrates the AUC value for different S .

As shown in Fig. 12, the smaller the S , LSM result is fractured with more details. In contrast, setting large S results in generating smooth and generalized landslide susceptibility map. LSM result with $S = 25$ strikes the balance between LSM result for small S and large S . According to Fig. 13, with the increment of S , AUC value first increases. After S reaches 25, AUC decreases with the increasing S . CPCNN yield the highest

AUC value (0.946) when $S = 25$, which explains why S was set as 25 in Section III-D. The cause of the above phenomenon is CPCNN with different size of input data have different emphasis on feature extraction. CPCNN with large size input data extract high-level features at a holistic level, while CPCNN with small size input data extract high-level features from a local perspective. CPCNN with $S = 25$ effectively extracts high-level features on large-scale and small-scale properly, and therefore performs the best.

D. Quality Level of Model Evaluation

In this study, the degree of model fitting (Training accuracy) and model prediction (Testing accuracy) were both evaluated. Based on the ranking criteria founded by Guzzetti [70], the quality of LSM model assessment in this research is on level V, which is relatively high compared with most LSM research. Fitting performance stands for the ability of the model to fit training data, while prediction performance measures the ability of model to predict the data not used in the training process. Theoretically, to obtain an unbiased evaluation result of prediction performance, training data should be completely independent from testing data. However, landslide inventory without temporal sequence is widely used in LSM research. For such inventory, no matter which strategy is used for training/testing data division, the correlation between training and testing data still exists. With the help of temporal landslide inventory, training and testing data can be divided according to temporal sequence, which could guarantee the independence of training and testing data. Thus, the evaluation results of this study could represent the unbiased predictive performance of CPCNN-ML.

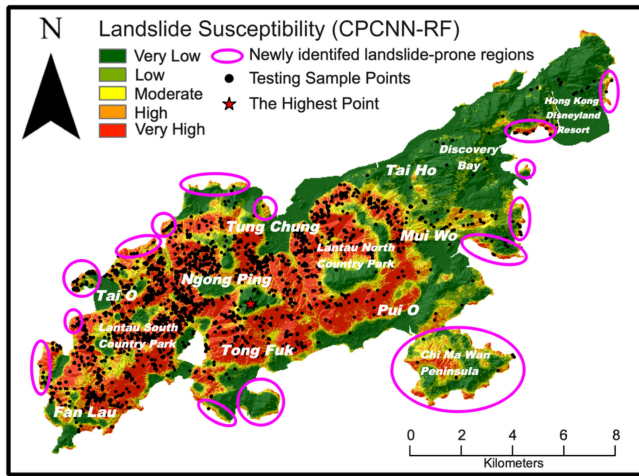


Fig. 14. Newly identified coastal landslide-prone regions.

E. Landslide-Prone Areas in Lantau Island

For all of the 10 landslide susceptibility maps in Figs. 9 and 10, overall distribution of landslide-prone areas is about the same. High and very high landslide susceptibility areas are mainly distributed in the southeast mountainous regions of Lantau Island. It is worth noting that more landslide-prone areas were marked by CPCNN-ML compared with traditional ML models. Taking CPCNN-RF as example (see Fig. 14), newly identified landslide-prone areas include the east and north coasts of Tai O, the north coast of Discovery Bay, the coast of Chi Ma Wan Peninsula, etc. It can be concluded that the proposed CPCNN-ML has a better ability in identifying coastal landslide-prone areas. Disaster prevention and mitigation department of HK should pay special attention to coastal regions marked in Fig. 14. Appropriate policy measures should be taken to reduce the impact of landslide-related geohazard on the corresponding areas.

V. CONCLUSION

Over the past decades, frequently occurred landslides in Lantau Island pose a great threat to the property and lives of HK citizens. In this study, by integrating modified PCNN model with tradition ML model, CPCNN-ML was proposed to map landslide susceptibility of Lantau Island. LSM results illustrate that landslides are highly likely to occur in the southeast mountainous and some coastal regions of Lantau Island. CPCNN-ML was compared with CPCNN, self-built CNN, and various traditional ML models. Based on experimental results and analysis, the conclusions can be drawn as follows.

- 1) CPCNN is superior to self-built CNN in LSM for its strong ability on deep feature extraction. By breaking through the limit on the number of channels of PCNN, CPCNN can be activated to fully extract high-level features from a multichannel predisposing factor matrix cube with the help of fine-tune strategy. By utilizing high-level features, CPCNN is capable of mapping landslide susceptibility with higher accuracy compared with self-built CNN. In

addition, the efficiency of CPCNN in the training process is prior to self-built CNN. What's more, CPCNN significantly reduces the number of hyper-parameters that needed to be optimized, which not only saves time greatly on hyper-parameter optimization, and more importantly, effectively avoids decline of prediction accuracy due to improper optimization.

- 2) CPCNN-ML paves a new way for LSM based on CNN for effectively fusing high-level features with numerical predisposing factors. High-level features extracted from predisposing factor matrix cube are tightly coupled with numerical predisposing factors, which enhance the diversity of feature type. LSM results generated by CPCNN-ML inherit the merits of CPCNN and ML models, which are fine-grained and reliable. Among four traditional ML models, RF appears to be the most effective model for CPCNN-ML. CPCNN-RF significantly improves the quality and accuracy of LSM through feature fusion and the combination of CNN and RF.

Given the above, the newly proposed CPCNN-ML is a promising model to predict landslide spatial occurrence with high precision. Achievements of this research can not only assist the HK government to formulate a disaster prevention and mitigation strategy, but also help to improve the precision and efficiency of landslide's early warning. However, it is worth noting that the selection of optimum hyper-parameter S is still based on trial-and-error. Therefore, in-depth research on the selection of appropriate S is of great necessity. Besides, limited by data source, only 10 predisposing factors were used in this research. And for the same reason, the CPCNN-ML model was only tested and evaluated on a district-scale, which failed to amply demonstrate its applicability. City-scale or country-scale study area and more types of predisposing factors should be considered in future research. Furthermore, a growing number of DL architectures are devised in recent years. Applying novel DL architectures in LSM research is another meaningful direction.

ACKNOWLEDGMENT

The authors thank GEO of CEDD for providing ENTLI dataset; the authors would like to thank editors and reviewers for their valuable comments.

REFERENCES

- [1] M. J. Froude, and D. N. Petley, "Global fatal landslide occurrence from 2004 to 2016," *Natural Hazards Earth System Sci.*, vol. 18, no. 8, pp. 2161–2181, 2018.
- [2] B.-G. Chae, H.-J. Park, F. Catani, A. Simoni, and M. Berti, "Landslide prediction, monitoring and early warning: A concise review of state-of-the-art," *Geosci. J.*, vol. 21, no. 6, pp. 1033–1070, 2017.
- [3] P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, and F. Guzzetti, "A review of statistically-based landslide susceptibility models," *Earth-Sci. Rev.*, vol. 180, pp. 60–91, 2018.
- [4] H. Zhang, G. Zhang, and Q. Jia, "Integration of analytical hierarchy process and landslide susceptibility index based landslide susceptibility assessment of the Pearl river delta area, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4239–4251, Nov. 2019.
- [5] F. Huang, J. Zhang, C. Zhou, Y. Wang, J. Huang, and L. Zhu, "A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction," *Landslides*, vol. 17, no. 1, pp. 217–229, 2020.

- [6] M. Scaioni, L. Longoni, V. Melillo, and M. Papini, "Remote sensing for landslide investigations: An overview of recent achievements and perspectives," *Remote Sens.*, vol. 6, no. 10, pp. 9600–9652, 2014.
- [7] F. Guzzetti, P. Reichenbach, M. Cardinali, M. Galli, and F. Ardizzone, "Probabilistic landslide hazard assessment at the basin scale," *Geomorphology*, vol. 72, no. 1–4, pp. 272–299, 2005.
- [8] G. G. Anagnostopoulos, S. Fatichi, and P. Burlando, "An advanced process-based distributed model for the investigation of rainfall-induced landslides: The effect of process representation and boundary conditions," *Water Resource Res.*, vol. 51, no. 9, pp. 7501–7523, 2015.
- [9] U. Haque *et al.*, "Fatal landslides in Europe," *Landslides*, vol. 13, no. 6, pp. 1545–1554, 2016.
- [10] C. J. Van Westen, E. Castellanos, and S. L. Kuriakose, "Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview," *Eng. Geol.*, vol. 102, no. 3–4, pp. 112–131, 2008.
- [11] V. Tofani, S. Segoni, A. Agostini, F. Catani, and N. Casagli, "Use of remote sensing for landslide studies in Europe," *Natural Hazards Earth Syst. Sci.*, vol. 13, no. 2, 2013.
- [12] D. Kanungo, M. Arora, S. Sarkar, and R. Gupta, "Landslide susceptibility zonation (LSZ) Mapping—A review," *J. South Asia Disaster Stud.*, vol. 2, no. 1, pp. 81–105, 2012.
- [13] F. Huang, C. Yao, W. Liu, Y. Li, and X. Liu, "Landslide susceptibility assessment in the Nantian area of China: A comparison of frequency ratio model and support vector machine," *Geomatics, Natural Hazards Risk*, vol. 9, no. 1, pp. 919–938, 2018.
- [14] S. Pirasteh, and J. Li, "Probabilistic frequency ratio (PFR) model for quality improvement of landslide susceptibility mapping from lidar-derived DEMs," *Geoenviron. Disasters*, vol. 4, no. 1, 2017, Art. no. 19.
- [15] J. Roy, S. Saha, A. Arabameri, T. Blaschke, and D. T. Bui, "A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2866.
- [16] Y. Achour, A. Boumezbeur, R. Hadji, A. Chouabbi, V. Cavaleiro, and E. A. Bendaoud, "Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria," *Arabian J. Geosci.*, vol. 10, no. 8, pp. 194, 2017.
- [17] F. Huang, Z. Cao, J. Guo, S.-H. Jiang, S. Li, and Z. Guo, "Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping," *CATENA*, vol. 191, 2020, Art. no. 104580.
- [18] C. Melchiorre, M. Matteucci, and J. Remondo, "Artificial neural networks and robustness analysis in landslide susceptibility zonation," in *Proc. IEEE Int. Joint Conf. Neural Netw. Proc.*, 2006, pp. 4375–4381.
- [19] P. Amatya, D. Kirschbaum, and T. Stanley, "Use of very high-resolution optical data for landslide mapping and susceptibility analysis along the Karnali highway, Nepal," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2284.
- [20] W. Chen *et al.*, "Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree," *Geocarto Int.*, vol. 34, no. 11, pp. 1177–1201, 2019.
- [21] X. Luo *et al.*, "Mine landslide susceptibility assessment using IVM, ANN and SVM models considering the contribution of affecting factors," *PloS One*, vol. 14, no. 4, pp. 1–18, 2019.
- [22] F. Huang, Z. Cao, S.-H. Jiang, C. Zhou, J. Huang, and Z. Guo, "Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model," *Landslides*, vol. 17, no. 12, pp. 2919–2930, 2020.
- [23] Y. Wang, H. Duan, and H. Hong, "A comparative study of composite kernels for landslide susceptibility mapping: A case study in Yongxin County, China," *Catena*, vol. 183, 2019, Art. no. 104217.
- [24] Y. Huang, and L. Zhao, "Review on landslide susceptibility mapping using support vector machines," *Catena*, vol. 165, pp. 520–529, 2018.
- [25] A. Arabameri, S. Saha, J. Roy, W. Chen, T. Blaschke, and D. T. Bui, "Landslide susceptibility evaluation and management using different machine learning methods in the Gallicash river watershed, Iran," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 475.
- [26] W. Chen *et al.*, "GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method," *Catena*, vol. 164, pp. 135–149, 2018.
- [27] Z. Fang, Y. Wang, G. Duan, and L. Peng, "Landslide susceptibility mapping using rotation forest ensemble technique with different decision trees in the three Gorges reservoir area, China," *Remote Sens.*, vol. 13, no. 2, 2021, Art. no. 238.
- [28] B. T. Pham *et al.*, "Landslide susceptibility assessment by novel hybrid machine learning algorithms," *Sustainability*, vol. 11, no. 16, 2019, Art. no. 4386.
- [29] F. Huang *et al.*, "Landslide susceptibility prediction considering regional soil erosion based on machine-learning models," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 6, pp. 377, 2020.
- [30] A. Shirzadi *et al.*, "A comparative study between popular statistical and machine learning methods for simulating volume of landslides," *Catena*, vol. 157, pp. 213–226, 2017.
- [31] Y. Zhang, T. Ge, W. Tian, and Y.-A. Liou, "Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China," *Remote Sens.*, vol. 11, no. 23, pp. 2801, 2019.
- [32] Q. Zhu, L. Chen, H. Hu, S. Pirasteh, H. Li, and X. Xie, "Unsupervised feature learning to improve transferability of landslide susceptibility representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3917–3930, Jul. 2020.
- [33] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [34] A. Dhillon and G. K. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, 2020.
- [35] A. B. Nassif, I. Shatin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, Feb. 2019.
- [36] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state-of-the-art," *Proc. IEEE*, vol. 105, pp. 1865–1883, Oct. 2017.
- [37] Y. Chen, D. Ming, and X. Lv, "Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation," *Earth Sci. Informat.*, vol. 12, no. 3, pp. 341–363, 2019.
- [38] K. Zhou, D. Ming, X. Lv, J. Fang, and M. Wang, "CNN-Based land cover classification combining stratified segmentation and fusion of point cloud and very high-spatial resolution remote sensing image data," *Remote Sens.*, vol. 11, no. 17, pp. 2065, 2019.
- [39] X. Lv, D. Ming, T. Lu, K. Zhou, M. Wang, and H. Bao, "A new method for region-based majority voting CNNs for very high resolution image classification," *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 1946.
- [40] N. Audebert, B. L. Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.
- [41] M. Brandmeier, and Y. Chen, "Lithological classification using multi-sensor data and convolutional neural networks," in *Proc. Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, 2019, pp. 55–59.
- [42] Y. Imamverdiyev, and L. Sukhostat, "Lithological facies classification using deep convolutional neural network," *J. Petroleum Sci. Eng.*, vol. 174, pp. 216–228, 2019.
- [43] W. Zhou, D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong, "SO-CNN based urban functional zone fine division with VHR remote sensing image," *Remote Sens. Environ.*, vol. 236, 2020, Art. no. 111458.
- [44] T. Lu, D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang, "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1496.
- [45] R. A. Marcum, C. H. Davis, G. J. Scott, and T. W. Nivin, "Rapid broad area search and detection of Chinese surface-to-air missile sites using deep convolutional neural networks," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042614.
- [46] Z. Shao, P. Tang, Z. Wang, N. Saleem, S. Yam, and C. Sommai, "BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 1050.
- [47] S. L. Uilo, M. S. Langenkamp, T. P. Oikarinen, M. P. DelRosso, A. Sebastianelli, and S. Sica, "Landslide geohazard assessment with convolutional neural networks using sentinel-2 imagery data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9646–9649.
- [48] O. Ghorbanzadeh, S. R. Meena, T. Blaschke, and J. Aryal, "UAV-based slope failure detection using deep-learning convolutional neural networks," *Remote Sens.*, vol. 11, no. 17, 2019, Art. no. 2046.
- [49] O. Ghorbanzadeh, and T. Blaschke, "Optimizing sample patches selection of CNN to improve the MIOU on landslide detection," in *Proc. 5th Int. Conf. Geograph. Inf. Syst. Theory, Appl. Manage.*, 2019, pp. 3–5.
- [50] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 196.

- [51] O. Ghorbanzadeh, D. Hölbling, S. R. Meena, and T. Blaschke, "Detecting earthquake-triggered large-scale landslides with different input window sizes convolutional neural networks," in *EGU General Assembly 2019-4477*, 2019.
- [52] Y. Wang, X. Wang, and J. Jian, "Remote sensing landslide recognition based on convolutional neural network," *Math. Problems Eng.*, vol. 2019, 2019, Art. no. 8389368.
- [53] Z. Lv, T. Liu, X. Kong, C. Shi, and J. A. Benediktsson, "Landslide inventory mapping with bitemporal aerial remote sensing images based on the dual-path full convolutional network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4575–4584, Mar. 2020.
- [54] C. Ye *et al.*, "Landslide detection of hyperspectral remote sensing data based on deep learning with constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5047–5060, Dec. 2019.
- [55] W. Shi, M. Zhang, H. Ke, X. Fang, Z. Zhan, and S. Chen, "Landslide recognition by deep convolutional neural network and change detection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3015826](https://doi.org/10.1109/TGRS.2020.3015826).
- [56] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 982–986, Jun. 2019.
- [57] Y. Wang, Z. Fang, and H. Hong, "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China," *Sci. Total Environ.*, vol. 666, pp. 975–993, 2019.
- [58] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping," *Comput. Geosci.*, vol. 139, 2020, Art. no. 104470.
- [59] V. D. Pham, Q.-H. Nguyen, H.-D. Nguyen, V.-M. Pham, and Q.-T. Bui, "Convolutional neural network—Optimized Moth flame algorithm for shallow landslide susceptible analysis," *IEEE Access*, vol. 8, pp. 32727–32736, 2020.
- [60] M.-F. J. Venture, "Final report on compilation of the enhanced natural terrain landslide inventory (ENTLI)," Maunsell-Fugro Joint Venture and Geotechnical Engineering Office, Hong Kong, 2007.
- [61] J. Nichol, and M. Wong, "Detection and interpretation of landslides using satellite images," *Land Degradation Develop.*, vol. 16, no. 3, pp. 243–255, 2005.
- [62] J. Huang, "Investigation on landslide susceptibility using remote sensing and GIS methods," 2014.
- [63] T. Stanley and D. B. Kirschbaum, "A heuristic approach to global landslide susceptibility mapping," *Natural Hazards*, vol. 87, no. 1, pp. 145–164, 2017.
- [64] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
- [65] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning (no. 2)*. Cambridge, MA: MIT Press, 2016.
- [66] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 111–118.
- [67] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 60, 2012, pp. 1097–1105.
- [68] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [69] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [70] F. Guzzetti, P. Reichenbach, F. Ardizzone, M. Cardinali, and M. Galli, "Estimating the quality of landslide susceptibility models," *Geomorphology*, vol. 81, no. 1–2, pp. 166–184, 2006.



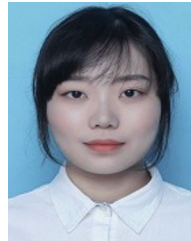
Yangyang Chen was born in Beijing, China, in 1992. He received the B.S. degree in geographic information science and M.E. degree in surveying and mapping engineering from China University of Geosciences, Beijing, China. He is currently working toward the Ph.D. degree at China University of Geosciences, Beijing, China.

His research interests include high spatial resolution remote sensing image processing and analysis, landslide information extraction, and semantic understanding of urban functional zone.



Dongping Ming received the B.E. degree in land administration and cadastral surveying from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1999, the M.E. degree in cartography and geographic information engineering from Wuhan University, Wuhan, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2006.

She is currently a Professor with the School of Information Engineering, China University of Geosciences, Beijing, China. Her research interests include remote-sensing image processing and analysis, information extraction from high spatial resolution satellite remote sensing image, and spatial statistics and its application.



Xiao Ling received the B.S. degree in geomatics engineering from Nanjing Normal University, Nanjing, China, in 2017. She is currently working toward the M.E. degree in surveying and mapping engineering in China University of Geosciences (Beijing).

She is currently working as a Post Graduate at the School of Information Engineering, China University of Geosciences, Beijing, China. Her research interests include landslide hazard and susceptibility analysis and high-resolution remote sensing image procession.



Xianwei Lv received the B.S. degree in geographic information science from East China University of Science and Technology, Nanchang, China, in 2016 and the M.S. degree in surveying and mapping from the China University of Geosciences, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

He does research in deep learning for very high-resolution image processing and applications.



Chenghu Zhou received the B.S. degree in land hydrology from Nanjing University, Nanjing, China, in 1984, and the M.S. and Ph.D. degrees in geographic information system from the Chinese Academy of Sciences (CAS), Beijing, China, in 1987 and 1992, respectively.

He is an Academician of the Chinese Academy of Sciences, Beijing, China, and an Academician of the International Eurasian Academy. His research interests include fundamental theory or application of remote sensing and geographical information system.