Unsupervised Feature Learning to Improve Transferability of Landslide Susceptibility Representations

Qing Zhu, Li Chen, Han Hu, Saeid Pirasteh, Haifeng Li, and Xiao Xie

***Abstract*—A landslide susceptibility map (LSM) is of vital importance for risk recognition and prevention. In the last decade, statistical methods have gradually exerted their impact on mapping the landslide susceptibility to locate the high-risk places of landslide. However, due to the complexity of getting full access to the thematic information in large scenarios, most of these statistical methods generally suffer from overfitting, inadequate representative power, and the inability to transfer the learned representation to other places. To solve these challenges, this study designed an unsupervised representation learning module, which features independence, compactness, robustness, and transferability. Specifically, we first stack restricted Boltzmann machines and denoising autoencoder to unsupervised discover the underlying representations embedded in the thematic maps. Then, we applied the transferring strategy in an adversarial manner to generalize the learned representations to the sample-scarce area. Experimental results and analyses using data in different regions have revealed that the proposed method can be generalized well between different LSM scenarios. In terms of precision, it outperforms other methods by a large margin, e.g., by around 7% compared to multilayer perceptrons with the same configuration, and by 3%–4% to the state of art algorithm random forest. Besides, compared to other methods, the landslide susceptibility map that is predicted by the proposed method featuring smoothness and stableness seems more reliable, and is more according to some prior knowledge that, for example, distance to the drainage, slope, and stratum, should exert dominant effects on the occurrence of a landslide.**

***Index Terms*—Denoised autoencoder (DAE), landslide susceptibility, restricted Boltzmann machines (RBMs), transferring learning, unsupervised representation learning.**

Manuscript received May 15, 2020; revised June 13, 2020; accepted June 27, 2020. Date of publication July 1, 2020; date of current version July 16, 2020. This work was supported in part by the National Natural Science Foundation of China under Project 41941019 and in part by the National Key Research and Development Program of China under Project 2018YFB0505404. *(Corresponding authors: Li Chen; Han Hu.)*

Qing Zhu is with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China, and also with the Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China (e-mail: zhuq66@263.net).

Li Chen, Han Hu, and Saeid Pirasteh are with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China (e-mail: browsercl@gmail.com; han.hu@swjtu.edu.cn; sapirasteh@swjtu.edu.cn).

Haifeng Li is with the School of Geosciences and Info-Physics, Central South University, Changsha 410006, China (e-mail: lihaifeng@csu.edu.cn).

Xiao Xie is with the Zhejiang Hi-target Space Information Technology Company Ltd., Huzhou 313200, China (e-mail: xiexiao@iae.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3006192

I. Introduction

THE Economic and Social Council (ECOSOC) of the United Nations adopted a resolution on July 2, 2018, and emphasized the quality of geospatial information and disaster risk reduction and management [1]. The landslide susceptibility map (LSM), which compromises the likelihood of occurrence of landslides for each location, is useful for comprehensive disaster prevention and mitigation [2]–[3]. It satisfies the natural hazards issue of the sustainable development goals (SDGs), which are conceived for a wide range of issues in local, national, regional, and global contexts. For the mapping of landslide susceptibility, related thematic maps and landslide inventory data are collected to produce trainable samples [4], [5]. With these samples, the purpose is to reliably predict the probability of other locations that are not sampled or even unseen. This study attempted the unsupervised representation learning (feature learning) [6] to improve the transferability of landslide susceptibility representations.

A. Objectives

In recent years, data-driven approaches [7]–[10] have arisen great interests with the development of remote sensing techniques that facilitates more thematic information. Despite the growing amount of works on the LSM, most of these methods have neglected the importance of utilizing the learned representation from a well-trained model, which causes them fail to transfer the learned knowledge to other datasets. Learning good representation from the input data is a core problem for unsupervised learning [11]. Although some recent studies [12] and [13] have shown good performance on extracting representation with impressive properties, there still exist some key problems needed to be fulfilled.

1) Small Labeled Samples: Although remote sensing data have become widely available and can be used for SDGs, the construction of landslide inventory generally needs expert knowledge or even on-site investigation [14]. Therefore, it can only be established in a certain region and the amount of samples in the landslide inventory is generally quite small. For supervised methods [15], [16], using small labeled samples will inevitably lead to the problem of overfitting. It may even lead to wrong models when the landslide inventory or the thematic maps contain noises. On the contrary, the remote sensing data or derived thematic maps are quite abundant. Thus, the intuitive idea is that

we can first learn useful intermediate representations from the thematic maps and apply such representation for the mapping of landslide susceptibility.

2) Limited Generalization Ability of Learned Model: Due to the difficulties during the investigation of landslide susceptibility inventories, it is generally only conducted in certain regions [17]. However, we can probably apply the LSM in other regions. The existing supervised approaches have rarely considered the issue of generalization [18], *i.e.,* learning a general model from one dataset that is suitably transferred to other (unseen) datasets. Especially when facing limited samples, the learned models from the existing models can only be used inductively. Therefore, a related issue is that, from a well-trained model using relatively large samples, can we transfer the learned knowledge to a specific region with small samples and improve the fitness of the model to that region [6]? This problem is not also discussed thoroughly in previous works.

B. Related Works

In this section, we first review current frequently used landslide susceptibility mapping methods, and then, introduce some new emerging unsupervised technologies that are beneficial to transferable representation learning.

1) Traditional Machine Learning: Physically-based feature engineering is too labor intensive to extract and organize discriminative information from the data [6]. At the same time, traditional machine learning methods show the preponderance for automatically finding the predominant factors and fitting the target classification function [19]. Those methods are explored and applied widely in landslide susceptibility mappings such as logistic regression [20], fuzzy logic [21], decision trees [22], support vector machine (SVM) [15], evidential belief function (EBF) [23], and weight of evidence [24]. Despite their good performance on the small scenario task, catching exact underlying discriminative factors is vulnerable to their defective representative ability to learn appropriate representations to fit the intractable objective, when facing problem with complex scenarios. The occurrence of a landslide is involved with many intertwined causative factors, thus requires a representative model to describe the correlation of these factors. To enhance such representative ability, Pham *et al.* [25] applied ensemble techniques that combine multiple base classifiers, which include Adaboost [16], bagging [26], dagging [27], and multiboost [28]. Nevertheless, in practice, the hyperparameters of these methods are generally elaborately determined according to empirical expertise, which could also be a labor-intensive process.

2) Supervised Learning: Aiming to alleviate the elaborative design of a discriminative representation, supervised learning LSM models are normally designed as an end-to-end nonlinear mapping from input to predictive susceptibility. The most common methods used in an LSM are decision-tree-related algorithms, such as the best-first decision tree [29], naïve Bayes tree [30], and random subspace [31]. Apart from these probabilistic graphical models, Wyner *et al.* [16] used a multilayer perceptron (MLP) and ensemble techniques. Besides, creative methods applying a convolutional neural network (CNN) begin to develop in recent years, *e.g.,* [32] and [33]. The representative power of a supervised learning method is proportional to the capacity of the model, with a fixed training manner [34]. When facing a classification task with a complex scenario, we may increase layer width and model depth to enhance the ability to map more complex nonlinearity to produce an end-to-end output. Correspondingly, more labeled data are required to fit the larger capacity model. Most supervised learning models are regarded as black box models for the illegibility of their internal learning manner, which leads to the biggest drawback of these models of their incapability of generalizing their learned representation to other dataset.

3) Unsupervised Transferable Representation Learning: Existing methods adapt in small scenarios well but neglect the designing of special features that could be transferred in different tasks. To achieve this goal, the unsupervised manner trains the model with unlabeled data to take a good insight into the data structure and learn a representation that disentangles more independent concepts to be encoded efficiently in a relatively low-dimensional space [35]. For example, Xu *et al.* [13] used the sparse autoencoder to reduce redundancy and correlation in the conditioning factors and applies regularized greedy forests (RGF) [12] to generate an LSM with sparse representation. Other works like PCA [36], isomap [37], and one-shot learning [38] skills are also used in landslide susceptibility mapping tasks. Recently, *DeepMind* have proposed a representation learning method that can learn representations from the various field using the same training mechanism—contrastive predictive coding [39] while guaranteeing the training efficiency, model robustness, and generalization ability. In many cases, test areas could suffer from the lack of valid labeled samples, but rare studies focus on transferring statistical strength and learning representations sharing the commonalities between a different dataset.

C. Contributions

In order to solve the aforementioned issues, the proposed method learns general and transferable representations in an unsupervised manner, to facilitate the learned model transferring from the sample-sufficient region to the sample-scarce region. Specifically, we first train stacked unsupervised learning modules [restricted Boltzmann machines (RBMs) and denoised autoencoder (DAE)] to learn such representations in the pretraining stage, and then, predict the LSM with a regressor using labeled data in the supervised training stage. Besides, to improve the transferability of learned representations, we apply an adversarial training method for extracting commonalities between different scenarios. For regions with no additional data, the learned model generalizes well. For the sample-scarce area, we could additionally transfer the knowledge from the well-trained model and adequate data of another region for better performances. We produced the sample set through GIS analysis on Arcmap and build the model in Python (version 3.5.6) under the Tensorflow framework.

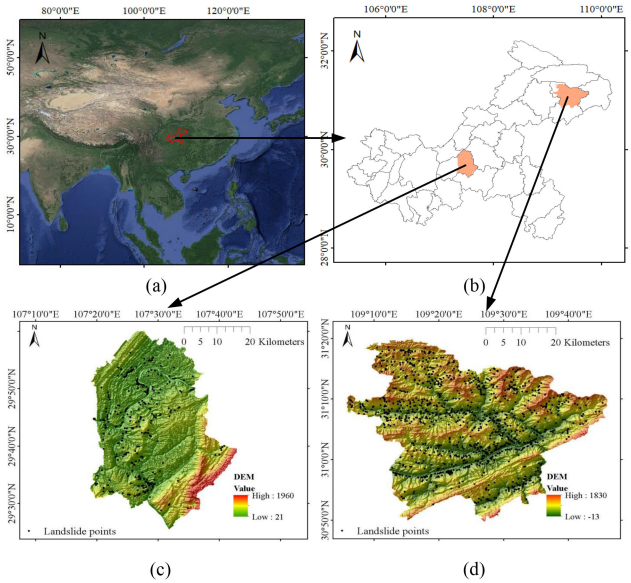


Fig. 1. Study area. (a) Geographic location of Chongqing. (b) Administrative map of Chongqing. (c) Fuling district. (d) Fengjie County.

In summary, the main contributions of the proposed method are as follow:

1) learning good representation with compactness, independence, robustness, and transferability;
2) using transferring approaches to apply well-learned knowledge between different areas.

The rest of this article is organized as follows. In Section II, we introduce the study area and dataset. Section III explains the proposed method. Experimental evaluations and discussions are presented in Section IV. Finally, Section V concludes this article.

II. Study Area and Dataset

We collected data from the two study area, Fengjie County (denoted as FJ) and Fuling District (denoted as FL), where recorded landslide examples in FJ are much more than that in FL. They are selected to evaluate the proposed method by transferring underlying knowledge from FJ to FL. Both of them are located in Chongqing, which is the largest city in China. Chongqing is featuring a mountainous area with severe topographic relief in southwest China, where frequent geological disasters cause a great loss of life and property. Among these occurred disasters, landslide account for a large proportion, e.g., the landslides in Three Gorges region, China [40], [41], and some typical landslide cases in FJ and FL, e.g. OuTang landslide, XinPu landslide, and TongMashu landslide. Fig. 1 gives an overview of the two study areas and the distributions of the landslide locations.

In this study, we collected the landslide inventory from Chongqing Geomatics and Remote Sensing Center. The inventory contains about 1000 locations of occurred and supposed to occur landslide locations in FJ and about 400 locations in FL. The potential susceptible landslide locations are determined from both the expert knowledge and on-site investigation. Besides, they are also considered as positive samples in this study. For each hazardous position, 17 factors are considered that might contribute to the occurrence of a landslide. The impacts of these factors on occurrence of a landslide are represented by the thematic information. They are consisting of rank of susceptibility, land use, stratum, digital elevation model (DEM), aspect, slope, curvature, normalized difference vegetation index (NDVI), sand distribution, clay distribution, silt distribution, vegetation, soil erosion, topographic wetness index (TWI), stream power index (SPI), distance to drainage, and distance to road. The thematic maps of FJ and FL are shown in Figs. 2 and 3, respectively.

III. Methods

A. Overview

The adopted architecture contains the generative part and the discriminative part, as is shown in Fig. 4. The former learns an intermediate representation [6] from the input features, which is more suitable for the regression tasks; for example, the learned representation is suitable for few-shot learning [42] and knowledge transferring [43]. The latter predicts the landslide susceptibility from the learned representation using the labeled samples.

1) Generative Part of the Architecture: The generative part contains a series of sequentially stacked unsupervised learning modules, which consists of two RBMs layer and a DAE layer. And the discriminative part is a full connection (FC) layer for predicting the probability. It should be noted that the generative part is pretrained unsupervisedly from only the unlabeled samples, by the strategy of greedy layerwise pretraining [19]. Specifically, RBMs are first trained with unlabeled data similar to deepBM [44] to disentangle the intertwined input features obtained from different thematic maps; the transformed outcome (upmost representation of the RBMs) is then used as the input to DAE, which aims to make the model noise-proof and decreases the model capacity by dimension reduction. Details of RBM and DAE are illustrated in Sections III-B and III-C, respectively.

2) Discriminative Part of the Architecture: For the discriminative part, the learned representation is further injected to the regressor, which outputs the probability of the occurrence of the landslide. We defines cross entropy as the objective function of the discriminative model, which measures the proximity between the training labels and model distribution, and predict the likelihood of a landslide case with *softmax* output. This part is learned by labeled samples after pretraining using the generative part.

3) Representation Transferring: Another essential property of the learned representation from the generative part is that it allows transferring learning. Because the learned representations decouple the correlation between different input factors, they make the model more suitable for transferring learned knowledge between different tasks. When applying the learned model from a large dataset to another region, they could effectively be fine-tuned and adapted to the unseen data, rather than training from scratch using only sparse samples. In addition, we also introduce a novel adversarial-based method to further increase the generalization ability of the model, during the fine tuning. It is specified in Section III-D.

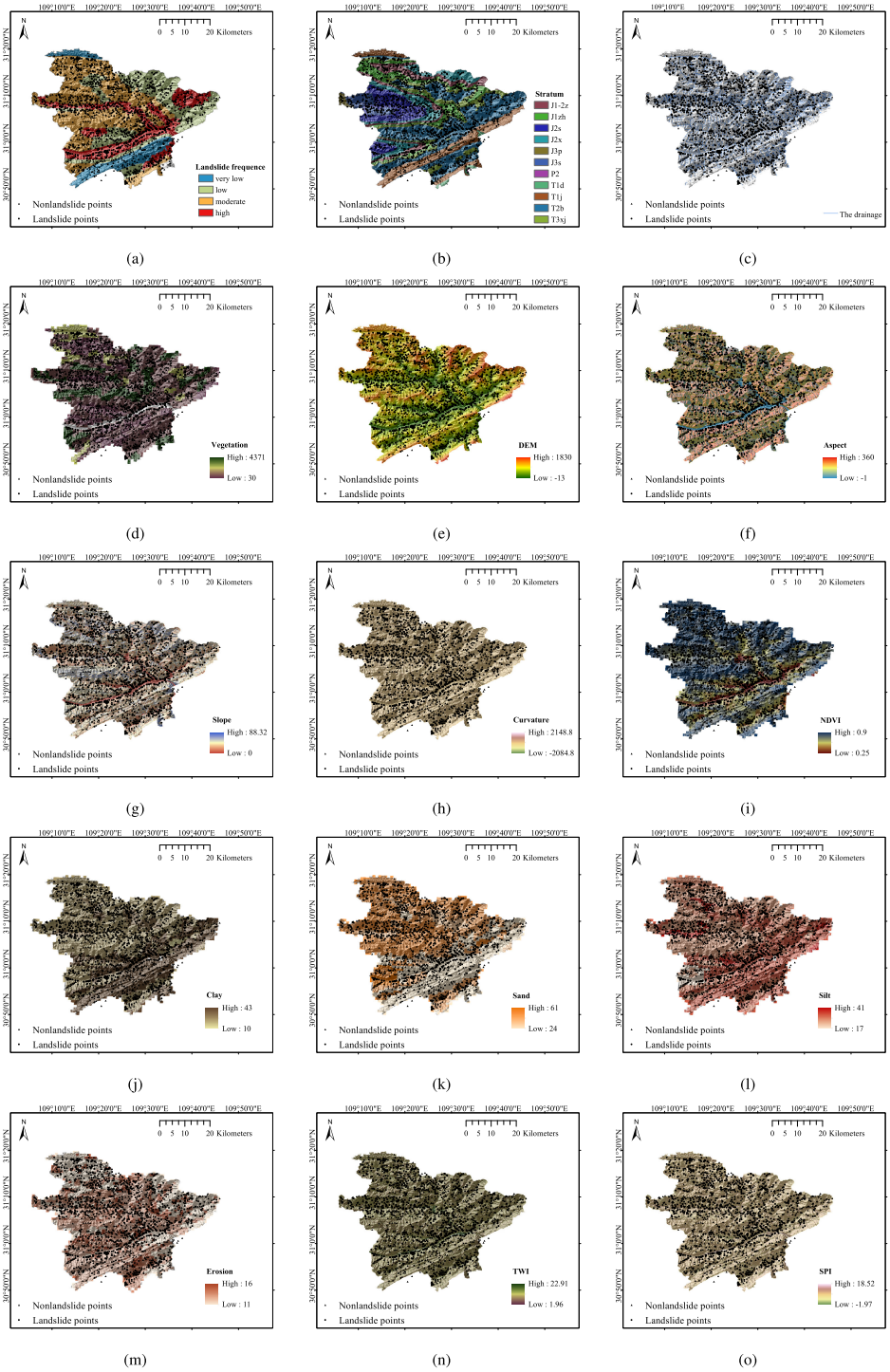


Fig. 2. Thematic maps in FJ used for extracting corresponding causative factors. (a) Landslide frequency. (b) Stratum. (c) Drainage. (d) Vegetation. (e) DEM. (f) Aspect. (g) Slope. (h) Curvature. (i) NDVI. (j) Clay. (k) Sand. (l) Silt. (m) Erosion. (n) TWI. (o) SPI.

B. Restricted Boltzmann Machine (RBM)

A restricted Boltzmann machine (RBM) is a stochastic neural network composed of two layers, which projects the input features into latent space. As shown in Fig. 5, m and n are the number of units in visible and hidden layers, respectively; $\mathbf{v}$ represents the state vector of input units, and $\mathbf{h}$ represents the state vector of inferred hidden units; a and b denotes the bias of visible units and hidden units, respectively, and W denotes the weight matrix. It is a generative model that can map the original input to an equivalent good representations, with designed superior properties.

During training, the updating rule of the parameters of RBM is demonstrated as follows:

$$\Delta W_{ij} = \Delta W_{ij} + \eta[P(h_i = 1|v^{(0)})v_j^{(0)} - P(h_i = 1|v^{(1)}v_j^{(1)})]$$

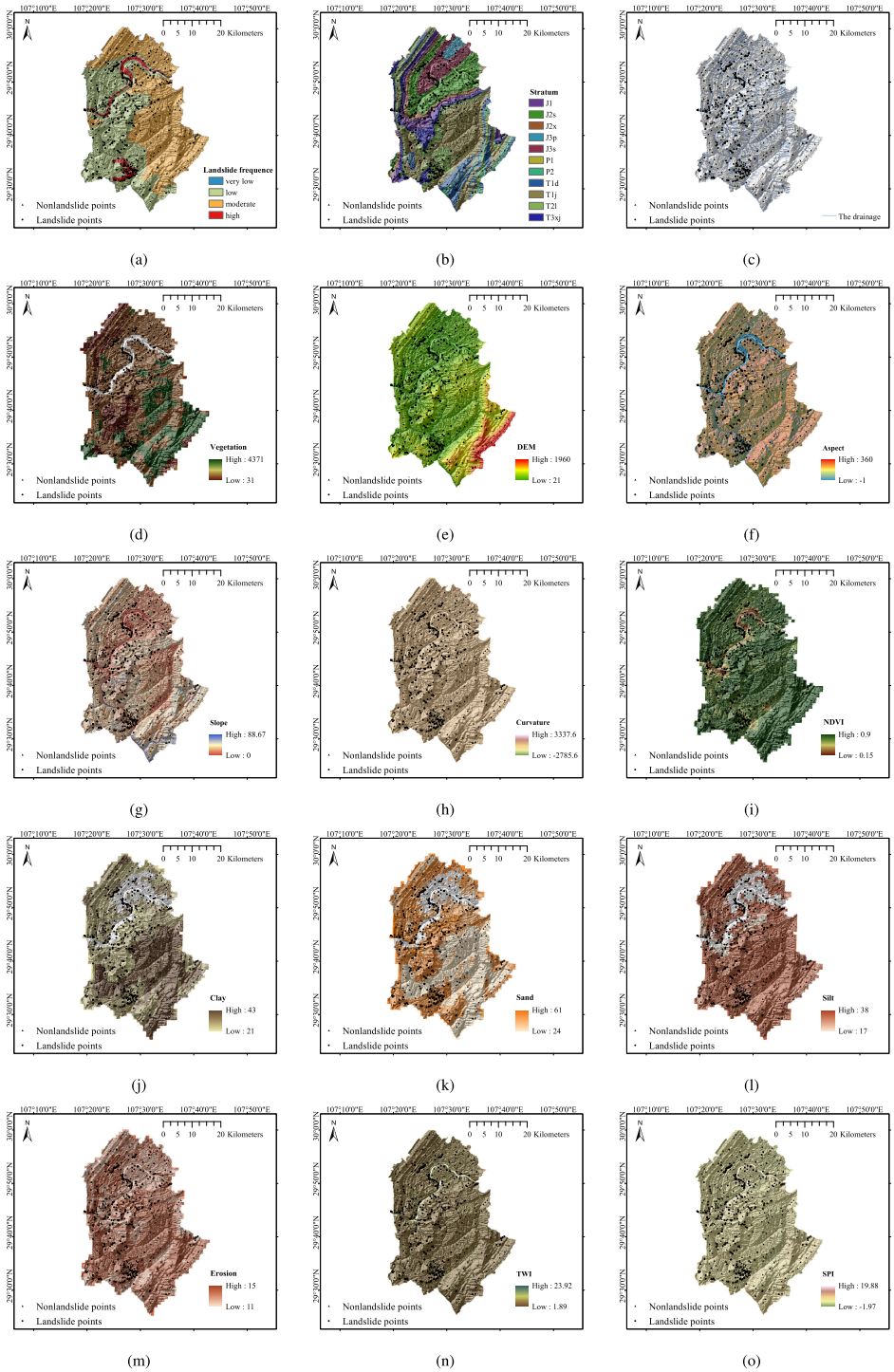


Fig. 3. Thematic maps in FL used for extracting corresponding causative factors. (a) Landslide frequency. (b) Stratum. (c) Drainage. (d) Vegetation. (e) DEM. (f) Aspect. (g) Slope. (h) Curvature. (i) NDVI. (j) Clay. (k) Sand. (l) Silt. (m) Erosion. (n) TWI. (o) SPI.

$$\Delta a_j = \Delta a_j + \eta[v_j^{(0)} - v_j^{(1)}]$$
$$\Delta b_i = \Delta b_i + \eta[P(h_i = 1|v^{(0)}) - P(h_i = 1|v^{(1)})] \quad (1)$$

where η denotes the learning rate, $v^{(0)}$ is the initial sample taken from the training set, $h^{(0)}$ is the inferred hidden state from $v^{(0)}$, $v^{(1)}$ denotes the reconstruction from $h^{(0)}$, and $v^{(k)}v_j$ represents kth Gibbs sampling state of the jth visible unit of the RBM.

Generally, the causative factors relating to thematic maps are correlated to describe a landslide. In the metric space, where we regard each factor as an independent dimension, the intertwined factors in the high dimension space could make it difficult for robustly fitting intricate objective. To disentangle such relationship, we utilize innate premise of the RBM that the hidden units are mutually conditionally independent basing on the mean field theory [45]. It transforms the input into hidden space, where each

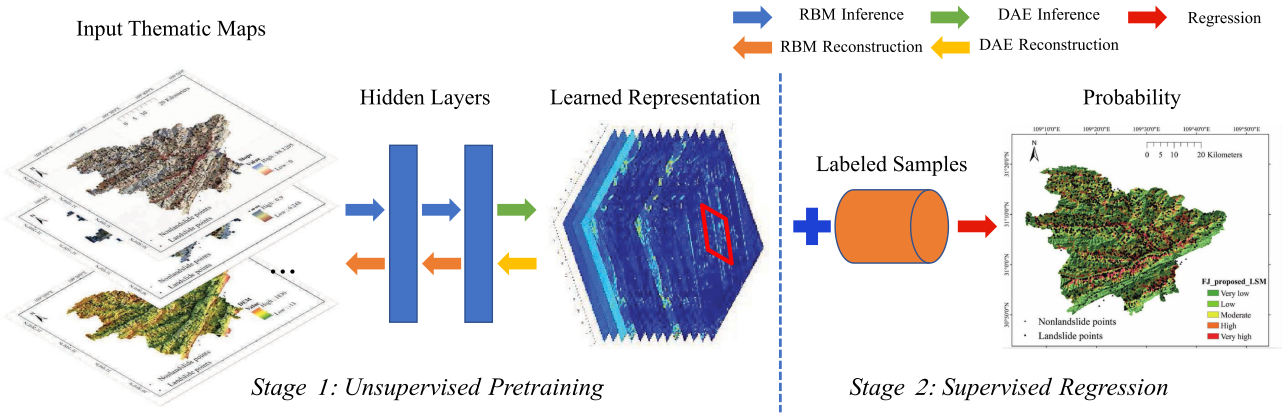


Fig. 4. Proposed architecture contains an unsupervised pretraining stage for extracting good intermediate representations with unlabeled samples, and a supervised regression process for predicting landslide susceptibility using such representation. The hidden layers are composed of a sequentially stacked unsupervised modules that contain two layers of RBMs with structure of $17 \times 32 \times 32$, and a DAE with structure of 32×16. While the regressor is in nature a fully connected (FC) layer with structure of 16×2.

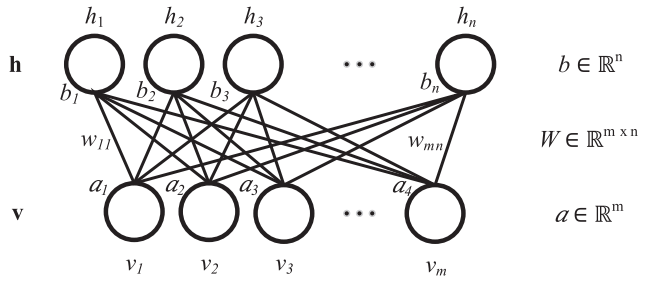


Fig. 5. Construction of the RBM. It introduces an energy-based pattern that describes the joint probability distribution of state **v**, **h**. The objective is to maximize the maximum likelihood of the given state. The derivatives of the training objective produces two intractable expectations, which can be inferred by the CD-k algorithm [19].

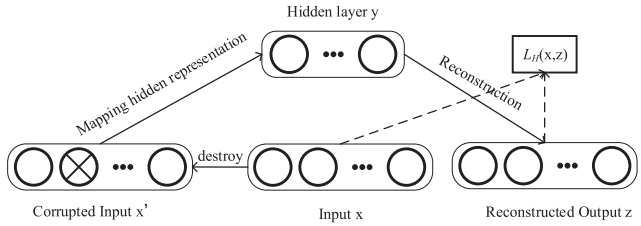


Fig. 6. Training process of a DAE. The input data x is first partially destroyed as x', which is then mapped to a hidden representation y. Finally, it reconstructs the input as z and trains the model with reconstruction error as the objective. $L_H(x, z)$ denotes the reconstruction loss between input x and reconstructed output z.

dimension is more likely to be orthogonal with other dimensions. Notice that we rise the dimension from 17 to 34 in hidden space for the purpose to amplify more independent and descriptive factors that share part of thought in 1×1 convolution of *network in network* [46].

C. Denoising Autoencoder (DE)

An autoencoder (AE) [47] network is a typical unsupervised learning module consisting of two parts, *i.e.*, encoder and decoder. The encoding process transfers the input feature to a hidden space, and the decoding process learns new representations that reconstruct the input data.

In this study, we use DAE [48], which is a variant of the AE layer. The training procedure of a DAE is shown in Fig. 6. The DAE is used to prevent noise perturbation of the data, by randomly setting 10% of units of the input features to 0, and then, minimizing the reconstructed loss. This is similar to the *dropout* mechanism [49] in deep learning. It also reconstructs the origin input space to a more compressed one.

For RBMs, the width of the output layer is not strictly determined. We used less but complete underlying factors to make up the new representation, by learning the reconstruction cost. In this study, we set the number of hidden units of *DAE* as 16.

D. Knowledge Transferring With Adversarial Training

The proposed transfer learning method exploits commonalities among various regions and transfers knowledge across different tasks [50]. During the transferring, the *source* model that is well trained is first directly copied to the well-trained *target* model. And then, we apply an adversarial-based method to urge *target* model to learn a representation that shares as many underlying factors as possible between source and target scenarios. It uses an MLP as the discriminator to judge whether the generative representation of a sample is from source or target distribution. The intuitive idea is that once the discriminator fails to correctly predict which domain (source and target) a sample comes from, the generated representation is thought to have caught enough commonalities between source and target data.

As shown in Fig. 7, adversarial training fine-tunes the initialized model in an adversarial manner. It forces a generator to produce representations that may confuse the discriminator, and in turn trains the discriminator to improve its identification ability of assigning correct domain label. The objective function

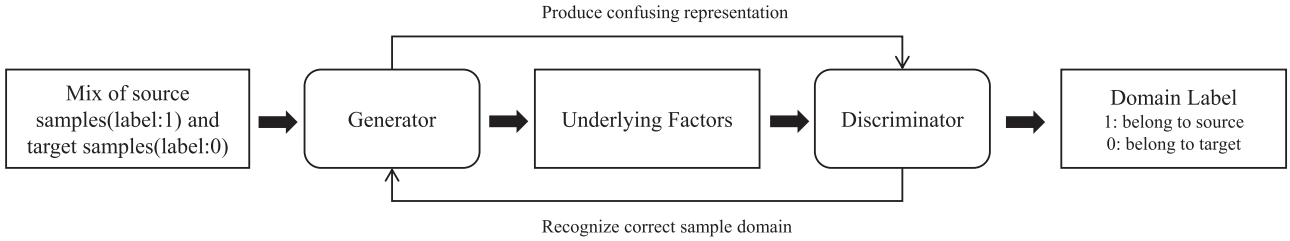


Fig. 7. Proposed adversarial training. For mixed sample set, we train a generator (generative part of proposed structure) to produce underlying factors that could confuse a discriminator, and in turn, we train the discriminator to predict the correct domain label.

F is given as (2):

$$\min_G \max_D F(D,G) = \mathbb{E}_{x_1:p_s(x)}[\log D(x_1)] + \mathbb{E}_{x_2:p_t(x)}[\log D(G(x_2))], \quad (2)$$

where D and G represent the discriminator and generator; $p_s(x)$ and $p_t(x)$ denote distribution of source data (FJ in this study) and target data (FL in this study); $D(x_1)$ is the probability of x_1 from the source sample rather than the target; $G(x_2)$ generates a representation from x_2 and $D(G(x_2))$ denotes the probability of the generative representation from the source samples. Training of D is to maximize the probability of assigning the correct domain label to input samples, while training of G is to maximize the probability of a transferred representation from x_2 being a source sample. It is equivalent to make both D and G use cross entropy of the predicted value and the label as their objective function. We alternately execute these two training process a few epochs until the generative model converges, *i.e.*, the generative model does not worsen the prediction performance of the discriminative model.

After learning the representation sharing commonalities between source and target data, we selects part of source samples to supplement the insufficient target samples. The current *target* model is then fine-tuned with the datasets augmented with generative samples drawn from the adversarial mechanism.

E. Implementation Details

The proposed method was implemented in Python (version 3.5.6) under the Tensorflow framework (tensorflow-gpu 1.5.0). The unsupervised modules and the regressor are all trained by Adam optimizer with one GTX1070 (memory 8 GB) that sets the batch size as 32 for each iteration. The learning rate of RBMs and DAE are, respectively, set to 1e-3 and 5e-5 and decreases by a factor of 0.02 each epoch. The appropriate training epochs of *RBM1*, *RBM2*, and *DAE* are, respectively, set as 30, 10, and 20, with the experimental verification of the reconstruction accuracy. For the FJ training set, the supervised training epoch of the whole model is set as 3000. For the FL training set, if not applied the transferring learning, the supervised training epoch is set as 1000, and 200 if applied for comparison. As for the adversarial training, the initial epochs of G and D are set as 500 and 200, and decreased by half each turn training (2). And we only use 30% FJ samples for *data supplementation* in the case that the FL model depends on FJ data a lot.

IV. Experiments and Results

A. Preprocessing

1) Valuation: For each landslide point, most of the 17 attribute values can be directly extracted from thematic maps through spatial analysis in GIS solutions, *i.e.,* Arcmap and QGIS [51], [52]. Specifically, the routine *spatial join* is used to connect different maps for landslide frequency, land use, and stratum; the routine *extraction values of points* is used for DEM, aspect, slope, curvature, NDVI, sand distribution, clay distribution, silt distribution, vegetation, soil erosion, TWI, and SPI; and the routine *proximity near* is used for distance to drainage and distance to the road. For different land use and stratum categories, we assigned 1–3 to represent their contribution to the occurrence of a landslide, according to [53], as is shown in Table I.

2) Validation Datasets: An unlabeled sample vector is made of the aforementioned acquired attribute values of a landslide point. Landslide samples are labeled as 1, while nonlandslide samples are labeled as 0. To avoid numerical problem, we normalized all dimension to [0,1] by

$$x^i_{\text{norm}} = x^i / x^i_{\text{max}} \quad (3)$$

where x^i_{norm} denotes the normalized value of the ith dimension, x^i denotes the value of the ith dimension in a sample vector, and x^i_{max} denotes the maximum of ith dimension.

There are about 1000 potential landslide samples in FJ and only less than 400 potential landslide samples in FL. Except for these positive samples, we randomly collected the same quantity of negative (nonlandslide) samples with preference based on some fundamental priors, e.g., low slope, near the drainage, and low landslide frequency. For both areas, we divide the sample vectors into training set (60%) and validation set (40%).

B. Evaluation and Comparison of LSM Models

To predict the landslide susceptibility map, we rasterized the FJ and FL area with a resolution of 120×120 m. For each raster cell, we produce the sample vector by means introduced in Section IV-A, and then, predict the likelihood by simply transposing the input vector to the *softmax* output.

1) LSM Prediction: Figs. 8 and 9 depict the LSM of FJ and FL with different approaches. There are five levels of LSM, including very low susceptibility with predicted value (0–0.3),

TABLE I
CONTRIBUTION OF LAND USE AND STRATUM

Land use										
Code	0100	0200	0300	0400	0500	0600	0700	0800	0900	1000
Contribution	2	2	2	2	1	3	2	3	3	3
Stratum										
Code	J1-2z	J1zh	J2s	J2x	J3p	J3s	P1	P2	T1d	T1j	T2b	T3xj
contribution	2	3	3	2	1	3	1	1	1	1	3	2

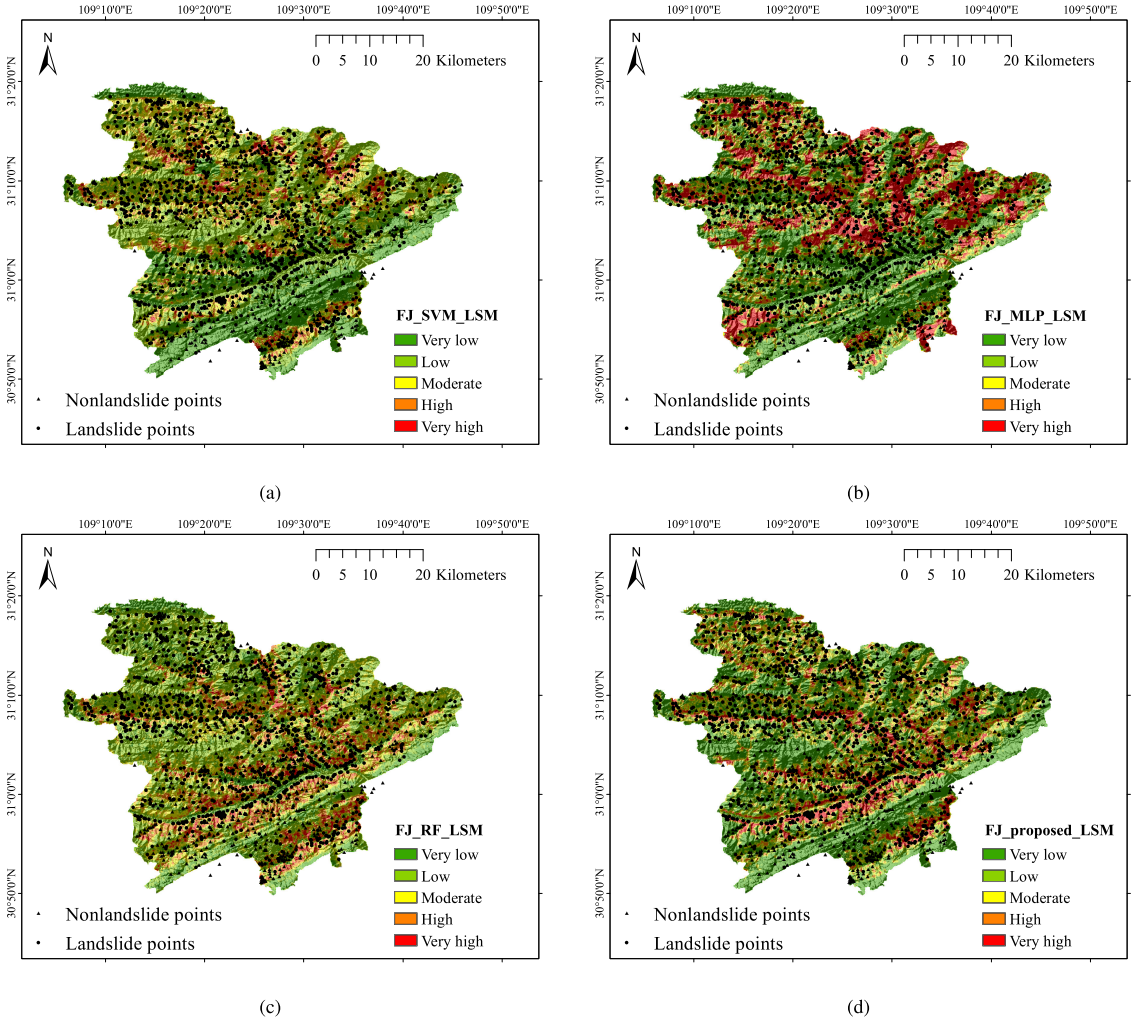


Fig. 8. Landslide susceptibility mapping of Fengjie County by (a) SVM, (b) MLP, (c) RF, and (d) proposed methods, respectively.

low susceptibility with predicted value (0.3–0.6), moderate susceptibility with predicted value (0.6–0.8), high susceptibility with predicted value (0.8–0.9), and very high susceptibility with predicted value (0.9–1.0). We find that the SVM-based and MLP-based method tend to give discontinuous patches, and fail to well simulate the distribution of landslide and nonlandslide points in some large local area. That is very likely to be a sign that the model is trapped in local minimum and suffers from overfitting. The random forest (RF)-based method gives a relatively better fitting for the distribution of these points, but also seems to be vulnerable to overfitting with a great fluctuant predictive value. In comparison with previous approaches, the proposed model gives the best fitting of landslide and nonlandslide distribution, with consecutive and stable LSM result.

We find the phenomena that could demonstrate our previous conclusion that, different from causative factors in original space, unsupervised modules learn a new representation where explanatory factors tend to change independently of each other while disentangling the factors of variation. We can easily find the distance to the drainage exerting a great influence on the prediction of the likelihood of a landslide, which illustrates the preferred landslide location of the landslide inventory. Likely, combined with Figs. 8 and 9 and Figs. 2 and 3, rank of susceptibility, stratum, slope, and NDVI can all be found having that correlation.

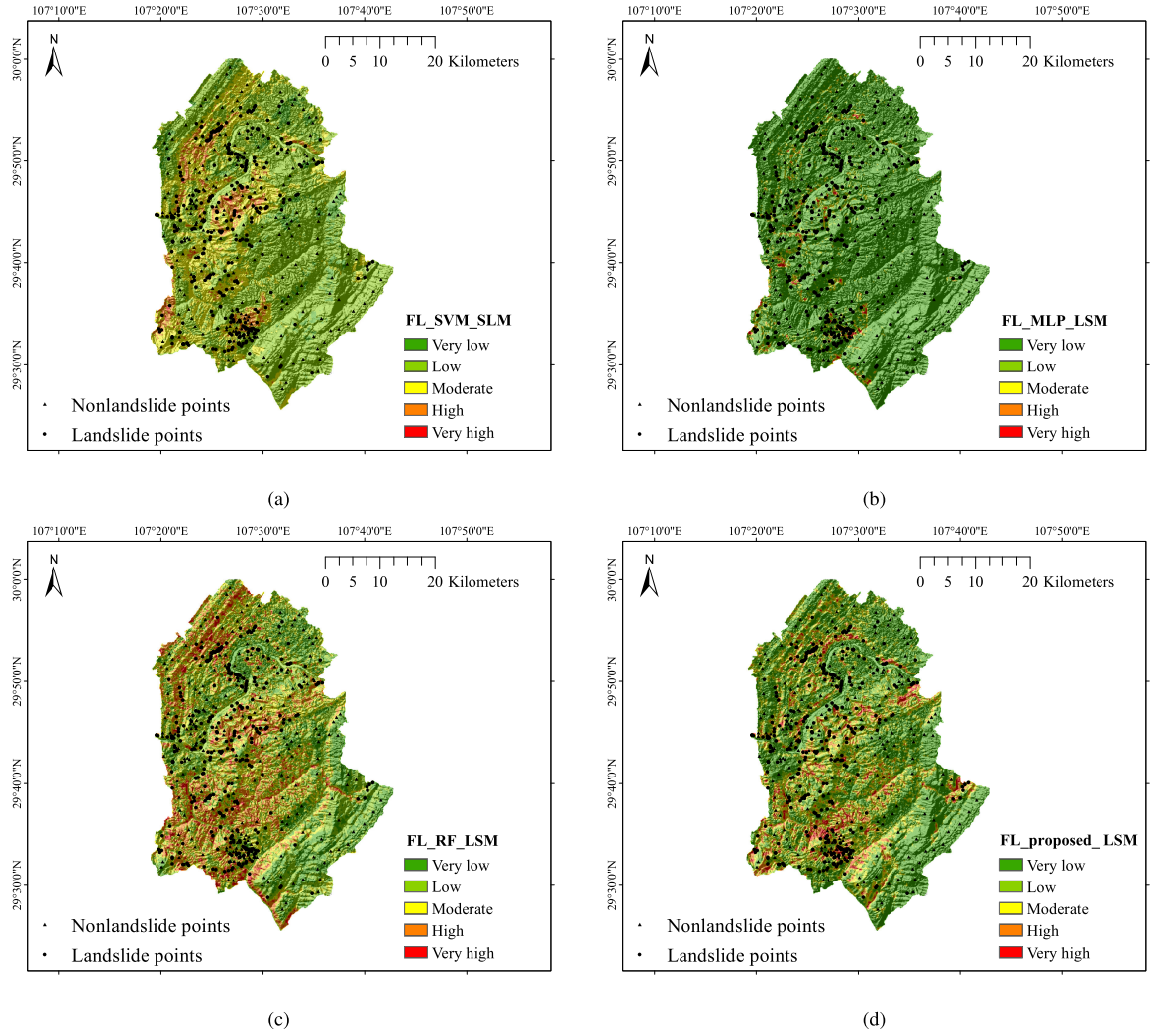


Fig. 9. Landslide susceptibility mapping of Fuling District by (a) SVM, (b) MLP, (c) RF, and (d) proposed methods, respectively.

2) Performance Evaluation: To evaluate the validity of the proposed method, we introduce statistical measures including area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, accuracy, precision, recall, and F1-score, as is shown in (4), where P is the number of landslide points, N is the number of nonlandslide points, and TP, FN, and FP denote true positive, false negative, and false positive, respectively. Specifically, TP means the number of landslide points correctly classified as landslide by the model; TN means the number of nonlandslide points correctly classified as nonlandslide; FP means the number of nonlandslide points wrongly classified as landslide; and FN means the number of landslide points wrongly classified as nonlandslide.

$$\begin{aligned}
\text{AUROC} &= \frac{\sum \text{TP} + \sum \text{TN}}{P+N} \\
\text{Sensitivity} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{Specificity} &= \frac{\text{TN}}{\text{TN}+\text{FP}} \\
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \\
\text{Precison} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\
\text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\
\text{F1-score} &= \frac{2\times\text{TP}}{2\times\text{TP}+\text{FP}+\text{FN}}.
\end{aligned} \tag{4}$$

In this study, we compared the predictive performance of the proposed method with SVM-, MLP-, and (RF-based approaches mainly in the following three aspects—transferability, stability, and efficiency. We use *sklearn lib* to draw the ROC curves of these models on the FJ and FL validation set, as is shown in Fig. 10. For each method, we shuffle the input data to get multiple (10) ROC curves and output the mean. The areas under the mean ROC curves (AUROC) are given in the lower right of the figure. In FJ with plentiful samples, the proposed method performs as good as RF, and slight advantages compared to SVM and MLP. The fact is that once given the adequate data quantity

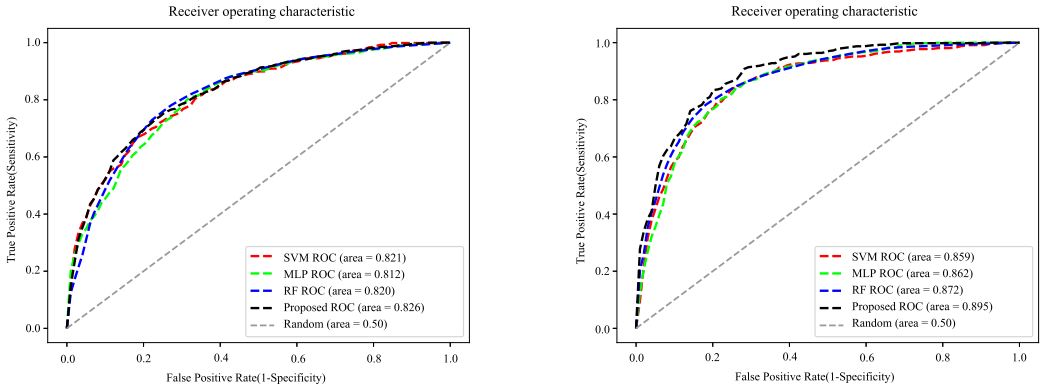


Fig. 10. ROC curves on validation set of (a) FJ and (b) FL.

TABLE II
LSM MODELS PERFORMANCE COMPARISON

Study areas	Models	AUROC	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
FJ	SVM	0.815	73.7	72.7	70.3	71.3
	MLP	0.810	73.5	72.3	69.9	70.9
	RF	0.816	73.8	**74.7**	65.3	69.4
	Proposed	**0.820**	**74.9**	70.2	**77.5**	**74.3**
FL	SVM	0.869	75.0	79.1	71.2	74.1
	MLP	0.855	76.7	77.1	76.6	76.6
	RF	0.862	77.8	**80.1**	73.3	76.3
	Proposed	**0.891**	**81.6**	78.7	**85.6**	**82.0**

TABLE III
TRANSFERRING ABILITY COMPARISON OF DIFFERENT LSM MODELS

Training set	Test set	Experiment design	SVM(%)	MLP(%)	RF(%)	Proposed(%)
FJ	FJ	*A*	73.7	73.5	73.8	**74.9**
FL	FL	*B*	75.0	76.7	77.8	**78.6**
FJ	FL	*C*	64.5	66.1	65.8	**75.2**
FL	FJ	*D*	59.2	60.4	61.1	**65.8**
FJ	FL	*E*	N/A	74.3	N/A	**80.5**
FJ & FL	FL	*F*	73.9	75.7	76.4	**77.8**
FJ & FL	FL	*G*	N/A	N/A	N/A	**81.6**

and quality, and most machine learning approaches could give a fairly well-fitting performance. But in FL with scarce samples, the proposed method shows its superiority by transferring unsupervised learned representations and sharing data strength. Despite that the discrepancies exist in each running of the ROC curves, the proposed method represents its preponderance and always get the best AUROC. In practice, we would frequently meet the circumstance where adequate data quantity and good data quality cannot be guaranteed. That is the core problem that can be largely solved by the proposed method.

The mean values of remainder evaluation statistics are also given in Table II. The bold entity means it performs the top score under a certain measure. These models have a floating performance due to the random initialization of the weight parameters and variational input data patches of each epoch. The proposed method tends to give the statistical measures with smaller "Stds" comparing to other methods, which demonstrates the proposed method a robust and a noise-proof algorithm.

C. Evaluation and Comparison of Transferring LSM Models

We transfer knowledge of the FJ to learn the FL model in a manner described in Section III-D. Within learned representation and supplemented dataset, the proposed method realizes substantial improvement in predictive performance, as is shown in Table II, while only trains current model with less than 200 epoch.

1) Evaluating Transferring Ability: To evaluate the superiorities of applied transferring skills, we implement some experiments as are shown in Table III, the left part explains the training and validation dataset, the right part represents the mean validation accuracies of referred approaches. Only experiments *A* and *B* do not apply the transferring skills; *C* denotes directly initializing the FL model with a well-trained FJ model, while *D* indicates the opposite. In *E*, the FL model is initialized with a well-trained FJ model and fine-tunes with FL labeled samples in few epochs at around 100, the training set of *F* is mixed with partial (30%) FJ and FL samples set. And in *G*, the adversarial mechanism is additionally applied to train the model with the mixed training dataset.

In the *A* and *B* experiments, the proposed method shows a slight surpassing performance. When transferring the weight value of FJ/FL to FL/FJ in *C*/*D*, we find the proposed method shows much better performance compared to other approaches. Except for the powerful generalization ability of the proposed

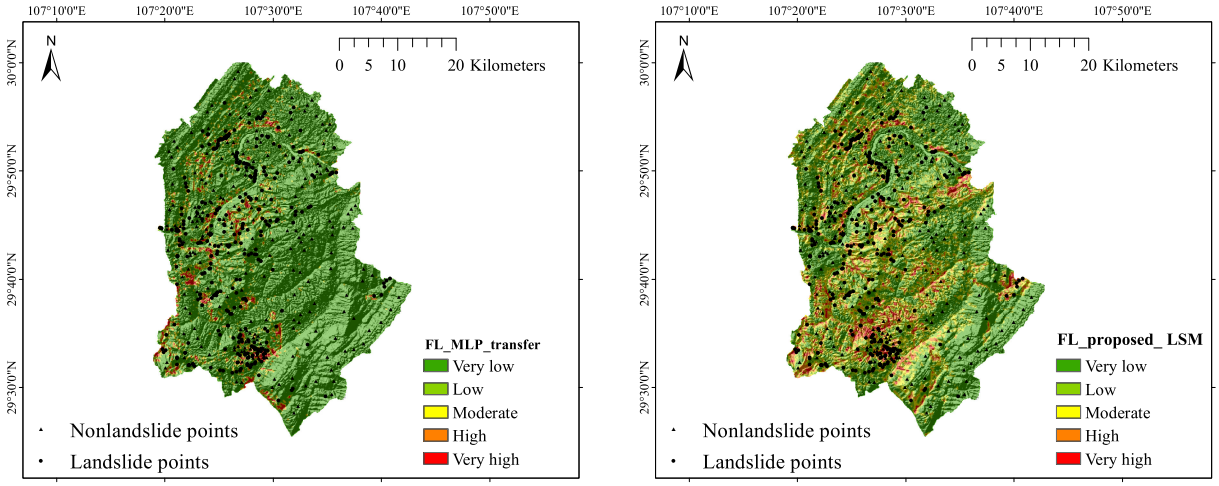


Fig. 11. Transferring performance comparison of the MLP and the proposed method in FL.

method, in *E*, it is found that the fine-tuning process is much more efficient and also can get stable test accuracy. Most importantly, to share data strength between FJ and FL, when applying the adversarial representation learning technique, we find a substantial improvement in validation performance. By judging the difference between train accuracy and test accuracy, we can determine whether a trainable model is suffered from overfitting. The SVM, MLP, and RF all suffer from overfitting with a difference value around 20%. Although RF is still yet the most popular algorithm for its top performance on predicting an LSM, its inability to transfer learned knowledge makes it limited to large scenario tasks. The proposed method also continuously gives a stable test accuracy within a small "Std" in all mentioned experiments.

2) Comparison to an MLP With Unified Configuration: To more fully evaluate the superiority of learned representations, we compare the proposed method with an MLP due to their structural similarity, and the prediction results are shown in Fig. 11. The hyperparameters of the two models are uniformed, e.g., layer width, depth, learning rate, optimizer, and batch size. The two models are initialized by training the FJ&FL set, and then, are finely tuned by FL labeled samples. The MLP roughly fits the training samples but fails to predict the susceptibilities in other regions where there are no samples. On the contrary, the proposed method not only better fits the training samples, but also it gives a better generalization performance, e.g., locating in a valley, close to the drainage, with a large slope, and in an affected stratum are predicted with relative high susceptibility. That demonstrates the successful transferring of the learned knowledge. The proposed model also shows a superior training efficiency in the FL task by a hybrid transferring skill, as is shown in Fig. 12. Whereas, transferring the MLP representation to a new scenario fails to facilitate the training efficiency and raise the prediction accuracy.

The FL model determines the likelihood of the occurrence of a landslide with underlying factors related to some dominant

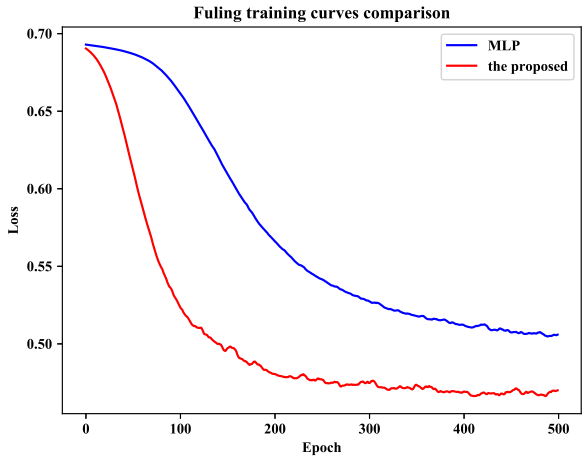


Fig. 12. Training efficiency comparison. Both the proposed model and the MLP are initialized by training FJ set with uniformed hyperparameters.

causative input similar to that of FJ, such as distance to the drainage, susceptibility, stratum, and slope. This illustrates the reason why the convergence to the optimum solution could be so fast—the FJ model has learned a representation within which many factors can be shared with the FL task. The advantage brought by applying this transferring technology is obvious. Within the representations learned from the source area, shared factors can be obtained and the target model parameters are initialized more close to the optimum, which facilitates the efficient and reliable prediction in the sample-scarce area.

D. Discussion and Limitation

This study utilized observed thematic information to produce LSM, which can support SDGs. Based on the aforementioned evaluation and analysis, we demonstrate the preponderance of the proposed method lies in good representation learning and

common knowledge transferring between multitasks. As is introduced in Section III, the RBMs is used to disentangle the intricate relationship of input causative factors and enhance the representation power of the model. While the DAE was used to make up a compressed and noise-proof representation by using less but complete underlying factors. The independence, compactness, and robustness of the learned representation prompt it more transferable from one dataset to the other. With the learned representation, we innovatively applied an adversarial training skill to urge the target area to share as much as commonalities with the source area, thus realizing data sharing and knowledge transferring. As compared to SVM, MLP, and RF, we experimentally evaluated the merits of the proposed method. It is mainly in the transferability of unsupervised learned features, the stability of statistical measure, and efficiency to train the model and predict the susceptibility. The proposed model also realize the highest generalized accuracy and AUROC in FJ with 0.749 and 0.820, and FL with 0.816 and 0.891.

We also summarized the current limitations of our proposed methods and gave the promising future work.

1) *Lack of utilizing image data:* Optical information is the primary data source of many LSM tasks, especially in CNN-based methods. The prospective research may focus on how to apply the deep level information of a CNN to constrain contributions of interferential nonlandslide factors.
2) *Nonautomatic selection of thematic information:* The principle to select thematic information in this study is to collect as much as the available information possibly contributes to a landslide. Although in upmost representation, the influence disentangled nonlandslide factors could be eliminated or weaken, which enlarges the model capacity and raises the training cost. Thus, it is helpful to automatically choose exact landslide causative factors before we consider a deep learning model.

V. Conclusion

The proposed LSM method applies remote sensing datasets for disaster risk reduction and response, which satisfies the natural hazards issue of the SDGs. In many scenarios, predicting the LSM by statistical methods normally suffers from insufficient labeled data. In the presented study, we stack unsupervised modules and train them to generate favorable representations that can facilitate and expedite the fitting of the intricate objective. A transferring method with adversarial training is then applied to transfer the common knowledge between different tasks. Based on the learned representations, the mentioned difficulties in Section I-A can be solved to a certain extent. In a large sample area, the proposed method tends to predict LSM featuring smoothness and stableness, meanwhile, according to prior knowledge of the occurrence of a landslide, which facilitates decision makers to make the right decision. While in a small sample area, the presented study provides developers or engineers with the idea to transfer learned underlying representations from a well-trained model. Thus, we can obtain a more reliable LSM predicting model.

References

[1] A. Rajabifard, *Sustainable Development Goals Connectivity Dilemma (Open Access): Land and Geospatial Information for Urban and Rural Resilience*. Boca Raton, FL, USA: CRC Press, 2019.

[2] L. Zhuo, Q. Dai, D. Han, N. Chen, B. Zhao, and M. Berti, "Evaluation of remotely sensed soil moisture for landslide hazard assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 162–173, Jan. 2019.

[3] S. Pirasteh and J. Li, "Landslides investigations from geoinformatics perspective: Quality, challenges, and recommendations," *Geomatics, Natural Hazards, Risk*, vol. 8, no. 2, pp. 448–465, 2017.

[4] W. Chen, S. Zhang, R. Li, and H. Shahabi, "Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling," *Sci. Total Environ.*, vol. 644, pp. 1006–1018, 2018.

[5] 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, p. 19, 2017.

[6] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[7] D. Van Dao *et al.*, "A spatially explicit deep learning neural network model for the prediction of landslide susceptibility," *Catena*, vol. 188, 2020, Art. no. 104451.

[8] A. Shirzadi *et al.*, "Shallow landslide susceptibility assessment using a novel hybrid intelligence approach," *Environmental Earth Sci.*, vol. 76, no. 2, p. 60, 2017.

[9] C. Ye *et al.*, "Landslide detection of hyperspectral remote sensing data based on deep learning with constrains," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5047–5060, Dec. 2019.

[10] 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.

[11] C. Zhang, Z. Yang, X. He, and L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," 2019. [Online]. Available: https://arxiv.org/abs/1911.03977

[12] M. I. Sameen, R. Sarkar, B. Pradhan, D. Drukpa, A. M. Alamri, and H.-J. Park, "Landslide spatial modelling using unsupervised factor optimisation and regularised greedy forests," *Comput. Geosci.*, vol. 134, 2020, Art. no. 104336.

[13] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[14] S. Pirasteh and J. Li, "Developing an algorithm for automated geometric analysis and classification of landslides incorporating LiDAR-derived DEM," *Environmental Earth Sci.*, vol. 77, no. 11, p. 414, 2018.

[15] J. Dou *et al.*, "An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan," *Natural Hazards*, vol. 78, no. 3, pp. 1749–1776, 2015.

[16] A. J. Wyner, M. Olson, J. Bleich, and D. Mease, "Explaining the success of Adaboost and random forests as interpolating classifiers," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 1558–1590, 2017.

[17] L. Li, R. Liu, S. Pirasteh, X. Chen, L. He, and J. Li, "A novel genetic algorithm for optimization of conditioning factors in shallow translational landslides and susceptibility mapping," *Arabian J. Geosci.*, vol. 10, no. 9, p. 209, 2017.

[18] B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro,"Exploring generalization in deep learning," in *Proc. Advances Neural Inf. Process. Syst.*, 2017, pp. 5947–5956.

[19] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[20] I. Colkesen, E. K. Sahin, and T. Kavzoglu, "Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression," *J. Afr. Earth Sci.*, vol. 118, pp. 53–64, 2016.

[21] T. Stanley and D. B. Kirschbaum, "A heuristic approach to global landslide susceptibility mapping," *Natural Hazards*, vol. 87, no. 1, pp. 145–164, 2017.

[22] W. Chen, X. Xie, J. Peng, J. Wang, Z. Duan, and H. Hong, "GIS-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-Bayes tree, and alternating decision tree models," *Geomatics, Natural Hazards Risk*, vol. 8, no. 2, pp. 950–973, 2017.

[23] H.-J. Oh, P. R. Kadavi, C.-W. Lee, and S. Lee, "Evaluation of landslide susceptibility mapping by evidential belief function, logistic regression and support vector machine models," *Geomatics, Natural Hazards Risk*, vol. 9, no. 1, pp. 1053–1070, 2018.
[24] Q. Ding, W. Chen, and H. Hong, "Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping," *Geocarto Int.*, vol. 32, no. 6, pp. 619–639, 2017.
[25] B. T. Pham, D. T. Bui, I. Prakash, and M. Dholakia, "Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS," *Catena*, vol. 149, pp. 52–63, 2017.
[26] B. T. Pham, D. T. Bui, and I. Prakash, "Bagging based support vector machines for spatial prediction of landslides," *Environmental Earth Sci.*, vol. 77, no. 4, p. 146, 2018.
[27] T. D. Acharya, "Regional scale landslide hazard assessment using machine learning methods in Nepal," Ph.D. dissertation, Kangwon Nat. Univ., Chuncheon, Korea, 2018.
[28] R. Jegadeeshwaran and V. Sugumaran, "Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features," *Measurement*, vol. 46, no. 9, pp. 3247–3260, 2013.
[29] R. Kohavi, "Scaling up the accuracy of Naive-Bayes classifiers: A decision-tree hybrid," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 1996, vol. 96, pp. 202–207.
[30] A. Liaw *et al.*, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[31] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
[32] V. D. Pham *et al.*, "Convolutional neural network-optimized moth flame algorithm for shallow landslide susceptible analysis," *IEEE Access*, vol. 8, pp. 32 727–32 736, 2020.
[33] 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.
[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[36] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, no. 1-3, pp. 37–52, 1987.
[37] O. C. Jenkins and M. J. Matarić, "A spatio-temporal extension to Isomap nonlinear dimension reduction," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 56.
[38] O. Vinyals *et al.*, "Matching networks for one shot learning," in *Proc. Advances Neural Inf. Process. Syst.*, 2016, pp. 3630–3638.
[39] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018. [Online]. Available: https://arxiv.org/abs/1807.03748
[40] C. Zhou *et al.*, "Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir Area, China," *Comput. Geosci.*, vol. 112, pp. 23–37, 2018.
[41] W. Wang and Y. Li, "Hazard degree assessment of landslide using set pair analysis method," *Natural Hazards*, vol. 60, no. 2, pp. 367–379, 2012.
[42] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Advances Neural Inf. Process. Syst.*, 2017, pp. 4077–4087.
[43] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," 2019. [Online]. Available: https://arxiv.org/abs/1911.02685
[44] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," in *Proc. Artif. Intell. Statist.*, 2009, pp. 448–455.
[45] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A, Statistical Mech. Appl.*, vol. 272, no. 1–2, pp. 173–187, 1999.
[46] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013. [Online]. Available: https://arxiv.org/abs/1312.4400
[47] A. Ng *et al.*, "Sparse autoencoder," *CS294A Lecture Notes*, vol. 72, no. 2011, pp. 1–19, 2011.
[48] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.
[49] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012. [Online]. Available: https://arxiv.org/abs/1207.0580
[50] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2018, pp. 270–279.
[51] J. Shaner and J. Wrightsell, *Editing in arcMap*, Esri, Redlands, CA, USA, 2000.
[52] QGIS Geographic Inf. Syst., QGIS Development Team, Open Source Geospatial Foundation Project, Beaverton, OR, USA, 2016.
[53] J. Qiao, C. Wu, and H. Tian, "Contribution rate research of stratum to landslide growth of Yunyang-Wushan segment in Three Gorges Reservoir region," *Chin. J. Rock Mech. Eng.*, vol. 23, no. 17, pp. 2920–2924, 2004.

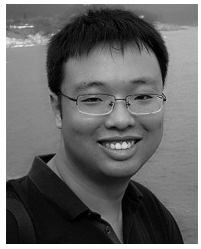
Qing Zhu received the B.S. and M.S. degrees in aerial photogrammetry from Southwest University, Chongqing, China, in 1986 and 1989, respectively, and the Ph.D. degree in railway engineering from North Jiaotong University, Beijing, China, in 1995.

From 1997, he had worked as a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, and received the Cheung Kong Scholars, in 2009. Since 2014, he has been a Professor and the Director of Research Committee with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. He also serves as Visiting Professor for the Wuhan University, Central South University, and Chongqing University. He has authored or coauthored more than 260 articles and seven books. His research interests include photogrammetry, geographic information system, and virtual geographic environment.

Dr. Zhu is the Editor-in-Chief for the *Journal of Smart Cities* and also serves as Editorial Board members of more than ten journals including *Computers, Environment and Urban Systems*, *Transactions in GIS*, and *International Journal of 3D Information*.

Li Chen received the B.S. degree from the School of Geology Engineering and Geomatics, Chang'an University, Xi'an, China, in 2016. He is currently working toward the Ph.D. degree with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China.

His research interests include machine/deep learning, photogrammetry, and geo-hazard remote sensing applications.



Han Hu received the B.S. degree from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2015.

He served as Postdoctoral Fellow with the Hong Kong Polytechnic University during 2015–2019. He is currently a Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. His research interests include photogrammetry, point clouds processing, and 3-D modeling.

Saeid Pirasteh received the Ph.D. degree in geology (remote sensing and GIS) from Aligarh Muslim University, Aligarh, India, in 2004, and also the Ph.D. degree in geography (GIS, geoanalytics, and LiDAR) from the University of Waterloo, Waterloo, ON, Canada, in 2018.

He is currently an Associate Professor with the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. His research interests include remote sensing and LiDAR data processing and applications in geology, environmental hazards, and disaster assessment toward implementing UN sustainable development goals 2030.

Dr. Pirasteh is the UN-GGIM Academic Network Member. He is also interested in integrating artificial intelligence, machine learning, computer vision, development of geospatial algorithms, models, software, mobile, and web app in geosciences applications.

Haifeng Li received the master's degree in transportation engineering from the South China University of Technology, Guangzhou, China, in 2005, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is currently a Professor with the School of Geosciences and Info-Physics, Central South University, Changsha, China. He was a Research Associate with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, in 2011, and a Visiting Scholar with the University of Illinois at Urbana-Champaign, Urbana, IL, USA, from 2013 to 2014. He has authored more than 30 journal papers. His current research interests include geo/remote sensing big data, machine/deep learning, and artificial/brain-inspired intelligence.

Dr. Li is a Reviewer for many journals.

Xiao Xie received the Ph.D. degree from Wuhan University, Wuhan, China, in 2016.

She is currently serving as a Senior Engineer with the Zhejiang Hi-target Space Information Technology Company Ltd., Huzhou, China. She is also an Assistant Professor of Urban and Environmental Computation with the Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China. Her research interests include 3-D geographic information system and smart cities.