Environmental Earth Sciences (2021) 80:239
https://doi.org/10.1007/s12665-021-09525-6

ORIGINAL ARTICLE

An approach based on socio-politically optimized neural computing network for predicting shallow landslide susceptibility at tropical areas

Viet-Ha Nhu[1] · Nhat-Duc Hoang[2,3] · Mahdis Amiri[4] · Tinh Thanh Bui[5] · Phuong Thao T. Ngo[6] · Pham Viet Hoa[7] · Pijush Samui[8] · Long Nguyen Thanh[9] · Tu Pham Quang[10] · Dieu Tien Bui[11]

Received: 13 June 2020 / Accepted: 26 February 2021 / Published online: 16 March 2021

Abstract
A new hybrid model approach based on Imperialist Competitive Algorithm, a socio-politically optimization, and neural computing networks (ICA-NeuralNet) was developed and proposed in this study with the aim is to improve the quality of the shallow landslide susceptibility assessment at the Ha Long city area, Quang Ninh province. This area, which belongs to one of the three key economic regions of Vietnam, has a high urbanization speed during the last ten years. However, the landslide has been a significant environmental hazard problem during the last five years due to extreme torrential rainstorms. For this regard, a geographic information system (GIS) database was established, which contains 170 landslide polygons that occurred during the last five years and ten influencing factors. The database was used for training and validating the ICA-NeuralNet model. The results showed that the integrated model achieves high performance with classification accuracy rates of 82.4% on the training dataset and 78.2% on the testing dataset. Therefore, the ICA-NeuralNet is subsequently employed for generating a landslide susceptibility map of the study area, which greatly supports the land-use planning as well as hazard mitigation/prevention of local authority.

Keywords Landslide · Imperialist Competitive Algorithm · Neural computing · GIS · Vietnam

Introduction

Recent research shows that climate change has played significant roles in inducing extreme rainfall events (Keellings and Hernández Ayala 2019; Liu et al. 2019; Sarhadi and Soulis 2017), which triggered a large number of landslides globally (Benz and Blum 2019; Haque et al. 2019). Consequently, landslides continue causing significant damages in many parts of the world, especially in mountainous and remote regions (Froude et al. 2018; Grahn and Jaldell 2017; Stäubli et al. 2018). Given their complex natures, it is still challenging to predict landslide occurrences, scales, and impacts with high accuracy. Thus, this fact limits the ability to reduce the risk of landslides and prevent or mitigate their devastating consequences, especially their direct effects on people's lives (Rossi et al. 2019). Moreover, in the context of increased population density and expansive infrastructure developments in mountainous areas, the risk and devastating consequences of landslides seem to be intensified. Therefore, innovative measure to reduce the adverse effects of landslides at the regional scale is an urgent need (Barraqué and Moatty 2019).

To reduce or even prevent damage caused by slope movements, it is necessary to develop models that allow identifying areas prone to landslides. Therefore, landslide susceptibility assessment is often regarded as an essential program for efficient spatial planning and risk management in mountainous areas (Cao et al. 2019; Ciurleo et al. 2019; Fressard et al. 2014; Kornejady et al. 2019; Yan et al. 2019; Zêzere et al. 2017). In recent years, GIS and statistical software tools have provided extensive opportunities for analyzing the spatial distribution of landslides (Chen et al. 2019b; He et al. 2019; Rossi et al. 2019). GIS technology has been increasingly implemented to evaluate cause and effect and investigate the impact of environmental changes on slope stability (Cascini 2008). It is because conventional physical models used to assess soil slope stability have significant obstacles,

✉ Viet-Ha Nhu
nhuvietha@humg.edu.vn

Extended author information available on the last page of the article

including the lack of access to crucial geotechnical information (Kuriakose and Beek 2009). GIS-based models, on the other hand, are quantitative methods for identifying sites with similar geological and geomorphological characteristics that rely on accessible remote-sensing data as well as field surveys with low-cost hand-held global positioning system devices (Guru et al. 2017).

The GIS data consisting of various layers of information (e.g., slope degree, aspect, topographic wetness index) provide a powerful tool for displaying and analyzing spatial patterns of landslide occurrences (Chen et al. 2018). The collected GIS data can be used to establish a landslide inventory map for a study area. The information stored in such an inventory map is subsequently extracted to create GIS datasets featuring landslide and non-landslide locations (Merghadi et al. 2018).

Various statistical methods have been employed to analyze GIS datasets for landslide susceptibility mapping at the regional scale (Huang and Zhao 2018; Pourghasemi et al. 2018; Reichenbach et al. 2018). Li et al. (2017) propose an improved version of the traditional frequency ratio method by enhancing the continuity of frequency ratio values and diminishing the subjectivity associated with the classifications of landslide affecting factors. Performances of the frequency ratio, weights of evidence, and statistical index models used for constructing multi-class weighted factors and generating landslide susceptibility maps have been investigated in (Razavizadeh et al. 2017); the authors conclude that the frequency ratio is highly sensitive to landslide influencing factors and all of the three approaches can deliver acceptable modeling accuracy for the study area. Ding et al. (2017) compare the capabilities of the frequency ratio, weights of evidence, and evidential belief function used for predicting landslide occurrence in mountain hinterland; the research finding is that the frequency ratio is simple yet effective method for computing landslide susceptibility maps. An approach that combines information theory and GIS has been put forward in (Tsangaratos et al. 2017); the information coefficients, approximated by Shannon's entropy index, are employed to specify the number of classes of each landslide influencing factors; accordingly, the evidential belief function and logistic regression are used for computing landslide susceptibility maps.

Based on the previous works' findings, the statistical methods are capable of delivering good prediction accuracy with a low computational expense. However, conventional statistical methods have certain limitations in analyzing complex, multivariate, and nonlinear data. The performances of such methods have been demonstrated to be inferior to advanced machine learning approaches (Merghadi et al. 2018; Tien Bui et al. 2019).

Therefore, machine learning approaches have been attracting researchers' attention to develop accurate landslide susceptibility prediction models in recent years (Achour and Pourghasemi2020; Mohan et al. 2020; Wang et al. 2020). Various machine learning methods have been used in recent years for the task of interest. They include neural network models (Pham et al. 2017a; Pradhan and Lee 2010), neuro-fuzzy models (Aghdam et al. 2017; Chen et al. 2019a; Sezer et al. 2011), support vector machines (SVM) (Hoang and Tien Bui, 2017; Pham et al. 2017b), alternating decision trees (Pham et al. 2017b), multivariate adaptive regression splines (Pourghasemi and Rahmati, 2018), radial basis function neural network (Pham et al. 2018a), decision trees (Pradhan 2013), classification and regression trees (Chen et al. 2017), and sophisticated ensemble learning approaches (Bandara et al. 2020; Kutlug Sahin and Colkesen, 2019; Sachdeva et al. 2020; Truong et al. 2018).

Nevertheless, comparative research works show that researchers of landslides at the regional scale have not yet reached a consensus on the most appropriate model for large-scale landslide susceptibility mapping (Huang and Zhao 2018; Reichenbach et al. 2018). It is because GIS-based landslide prediction is heavily data dependent; a specific machine learning method appropriate for one study area may not be so for another study area. Therefore, investigating other novel alternative machine learning methods is a worthy direction.

Among methods used for landslide spatial mapping, neural computing network (NeuralNet) remains one of the most widely employed machine learning models (Aditian et al. 2018; Harmouzi et al. 2019). The NeuralNet method possesses a strong capability in dealing with multivariate and nonlinear data (Pham et al. 2018b; Wang et al. 2016). Due to universal learning capability (Bishop 2011), a NeuralNet model can construct a complex decision boundary that accurately separates the dataset under study into two categories of landslide and non-landslide coupled with probabilistic outputs. These results can be subsequently used to establish a landslide susceptibility map for the study area.

With the recent development of machine learning, the traditional stochastic gradient descent and backpropagation algorithm from the 1990s have been largely ignored (Ma et al. 2019) due to unstable performance and sensitivity to starting conditions. Thus, they hinder the prediction performance of the NeuralNet in landslide modeling significantly. Accordingly, scholars and practitioners have resorted to metaheuristic methods as alternatives for training NeuralNet models. Various metaheuristic approaches, including Differential Evolution (Wang et al. 2015), Social Spider Optimization (Mirjalili et al. 2015), Genetic Algorithm (Göçken et al. 2016), Krill Herd Algorithm (Kowalski and Łukasik 2016), Particle Swarm Optimization (Chatterjee et al. 2017), Spotted Hyena Optimizer (Li et al. 2018), Ant Lion Optimization (Moayedi et al. 2019), Artificial Bee Colony (Ghaleini et al. 2019) and many others, have demonstrated the capability of

constructing NeuralNet models with good predictive accuracy. Therefore, applying metaheuristic in NeuralNet training becomes an increasing research trend (Ojha et al. 2017).

Besides, the Imperialist Competitive Algorithm (ICA) (Atashpaz-Gargari and Lucas 2007) is an attractive metaheuristic based on novel social–political concepts that can be used for optimizing neural computing models. Nevertheless, an ICA-optimized neural computing model's performance has rarely been explored in shallow landslide assessment. Thus, this work attempts to fill this gap in the current literature by proposing ICA-NeuralNet for landslide susceptibility mapping, with a case study at a tropical area of the Ha Long city and the Cam Pha city, which belong to the northeastern mountainous region of Vietnam. Finally, conclusions are given.

Background of the employed computational methods

The overall research method used in this study comprises five main steps, as shown in Fig. 1. In the first and second phases, the landslide inventory and the GIS database for the study area are established. The GIS database, containing information on landslide occurrences and influencing factors, is then used to train the proposed ICA-NeuralNet. The ICA-NeuralNet is a hybridization of neural computing techniques and metaheuristic optimization. After the proposed hybrid model is constructed, its predictive performance can be evaluated. Finally, the model is applied to produce a landslide susceptibility map for the whole study area.

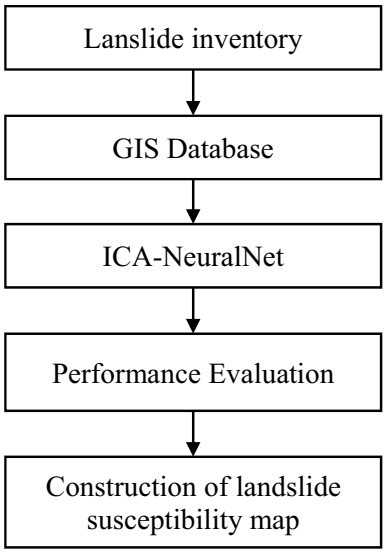

Fig. 1 The overall research method of the study

Neural computing network

Neural computing network (NeuralNet) is a supervised machine learning method that originates from actual biological neural networks (Pham et al. 2018b). As a supervised algorithm, a NeuralNet model used for shallow landslide susceptibility mapping can be trained with historical data stored in a landslide inventory. A landslide inventory must include a set of input features and ground truth labels ("landslide" and "non-landslide"). Due to its universal approximation, the NeuralNet is capable of mining the functional relationship between a set of input features and the ground truth labels.

Generally, a NeuralNet model contains three layers: the input, hidden, and output layers (Sundermeyer et al. 2012). The first layer receives information regarding the input features of shallow landslide occurrences. The hidden layers containing artificial neurons process the input features to derive the final class labels of shallow landslide occurrence (either "landslide" or "non-landslide"). The NeuralNet is able to learn complex concepts due to a sophisticated organization of interconnected neurons. Notably, the feasibility of capturing nonlinear mapping functions in NeuralNet models is achievable through the employment of nonlinear activation functions used by learning units (called neurons) in the hidden layers. The final layers compute the output class probability of "landslide" and "non-landslide".

It is noted that an NN model can adapt its structure through a process of fine-tuning its connecting weights. Let us consider a landslide data X with D influencing factors $f : X \in R^D \rightarrow T \in R^c$ be a mapping function that computes the class output T with c output classes, based on information provided by X, which are a set of real numbers (R). In this research context, c is equal to 2, which are the two classes: landslide and non-landslide, and. The final output of a NeuralNet model used for shallow landslide susceptibility mapping is compactly stated as follows (Pham et al. 2018b):

$$f(X) = b_O + OW \times (f_A(b_I + IW \times X)), \quad (1)$$

where IW and OW represent weight matrices of the hidden layer and the output layer, respectively. $b_I = [b_{11}, b_{12\ldots}, b_{1N}]$ denotes a bias vector of the hidden layer; b_O denotes a bias vector of the output layer; f_A represents an activation function (e.g., log-sigmoid). Herein, $C = 2$ denoting the number of class labels. It is also worth noticing that the weight matrices (IW and OW) and the bias vectors (b_1 and b_2) of the NeuralNet can be computed directly from data with ground truth labels using the error backpropagation framework (Hegazy et al. 1994; Rumelhart et al. 1986).

Imperialist competitive algorithm (ICA)

The Imperialist Competitive Algorithm (ICA) is a metaheuristic method inspired by the behaviors of the human's socio-political community. This metaheuristic algorithm imitates the real-world social policy of imperialism within which empires rise and fall. Notably, a rising empire starts to dominate colonies and harness their sources. Various empires, represented by solution candidates in the ICA population, compete to gain supreme domination. Throughout the competition process, an empire may lose colonies and decline. For further demonstration of this algorithm, users are guided to the previous works of Hosseini and Al Khaled (2014), Le et al. (2018), and Ding et al. (2019).

The algorithm's main stages include the formation of early empires, assimilation, revolution, and imperialist competition (Atashpaz-Gargari and Lucas 2007). The ICA has significant features such as good neighborhood search capability, effective global search property, and good convergence speed (Hosseini and Al Khaled 2014). The features of the ICA allow its searching process to have a good chance to escape from local optima and prominent structure.

In recent years, ICA has been used to solve various complex optimization problems, revealing excellent features in terms of convergence and global search properties (Gerist and Maheri 2019; Mikaeil et al. 2018; Tashayo et al. 2019; Tien Bui et al. 2018; Wang et al. 2019). In this study, the ICA is used to optimize the NeuralNet model structure employed for shallow landslide susceptibility mapping.

The general ICA procedure is summarized as follows (Atashpaz-Gargari and Lucas 2007):

1. Select some random points from the performance and place the empires in the initial stage,
2. Attraction: it makes the colonies of each empire closer to the imperialist,
3. Revolution: randomly change the position of some countries,
4. If a colony empire has a lower cost colony, change the colonial and imperialist positions.
5. Unite the same empires,
6. Calculate the total cost of empires,
7. Select the weakest colony (colonies) from the weakest empires and give it (one) to one of the empires (imperialist competition),
8. Destroy the powerless empires,
9. If the stop condition is met, then stop.

Study area and data

Description of the study areas

The study area belongs to Quang Ninh province, located in the north of Vietnam, covering a region of approximately 563 km^2. The study area is located between the latitudes 20° 40′ 00″ and 21° 13′ 00 " N and between the longitudes 106° 55′ 00″ and 107° 25′ 00″ E. The region has a tropical climate where the monthly average temperature is 24 °C, and the annual precipitation is 2306 mm (Pham 2018).

The geological evolution and tectonic movement in the study area have resulted in faults, folds, trenches, terrains, basins, and more than ten formations (Thanh 2012), which generate a diversity of hilly and mountainous terrains today. Herein, the terrains arranged and extended in an east–west direction coinciding with the fold axis direction fluctuate from 0 m to 829.1 m above sea level with slope angles varying from 0° to 76.73°.

The geodynamics relates mainly to the two major fault zones, the Red River and the Tan-Lu. They strongly link to Himalayan and Indosinian orogenic cycles (Fenart et al. 1999). The first zone is the collision result of the Asia plate and the India plate, which divides the Indochina block and the South China Block (Fenart et al. 1999). The second one is a large dynamic tectonics belt, which generates a sliding ramp stretching from the Halong region (the study area) to the Hong Kong region (Zhang et al. 2015).

Together with the geodynamics activities, the Quang Ninh province often suffers from prolonged and heavy rains (Loi et al. 2017), especially from 2015 to now. For example, the most massive rainfalls in the past 40 years (> 600 mm) during the four days from 25 to July 28, 2015, triggered various landslides and flash floods in the province, causing 17 people died, six people to missing, many residential areas were submerged. The total estimated damage is more than 40 million USD (Quyen 2015). In the study area, landslides occurred in all mountainous areas, especially in areas with a high density of rivers and streams, thin vegetation cover, high population density (Ha et al. 2020; Loi et al. 2017; Nguyen et al. 2020; Nhu et al. 2020; Van et al. 2017).

Shallow landslide inventory

It is noted that this study relies on NeuralNet and ICA to construct a machine learning-based model used for spatial analysis of landslide occurrence. One fundamental assumption of landslide susceptibility prediction is that the factors causing past landslides will continue to influence the likelihood of landslide occurrence in the future (Reichenbach et al. 2018). Hence, it is required to collect geo-information on past landslides such as terrain, geological condition, and

land-use. They are used as the basic data for establishing a shallow landslide inventory in this study. In this analysis, a landslide inventory with 196 shallow soil and rock mixed soil slides from the State-Funded Landslide Project No.105.08–2017.316 of Vietnam (Nhu et al. 2020) was used. The locations of landslide occurrences were determined using one-meter resolution aerial photographs and Google Earth images 2019. This helps to correct landslide locations, which is a good database for modeling. Herein, the landslide scars concept (Nefeslioglu et al. 2008) was used to map these failure features. Besides, we only consider only rainfall-triggered landslides because no earthquake-triggered landslide was reported in the study area.

These landslides occurred in the last 5 years, from 2015 to 2019. Figure 2 shows some photos of shallow landslides in the study areas. Our fieldwork and investigations showed that during and after those heavy rains, these soils and rock mixed soils were saturated with water, and therefore, leading to landslides. Besides the weak soil characteristics of the study areas, the thin vegetation cover makes the process of saturation of the rock mixed soil layer on the surface occurs very quickly. Thus, long-term rainfall intensity combining with the complex geological structure, various landslides in the study areas have also been activated.

The constructed inventory map (Fig. 3) has identified the total number of 3730 pixels of landslide occurrences consisting of 1865 pixels of landslide and 1865 pixels of non-landslide are randomly sampled from the map of the study area. In order to train the model, the data of landslide locations are randomly divided into Model Training and Model Validating in a 3:1 ratio. The numbers of samples in training and validating sets are 2722 and 1008, respectively.

Fig. 2 Landslide photos in the Ha Long (**a**) and Bai Chay (**b–d**) of the study area. The photos **a**, **b**, and **c** were taken in August 2015 (photos courtesy of baoquangninh.com.vn), whereas photo (**d**) was taken by Viet Ha Nhu in July 2019

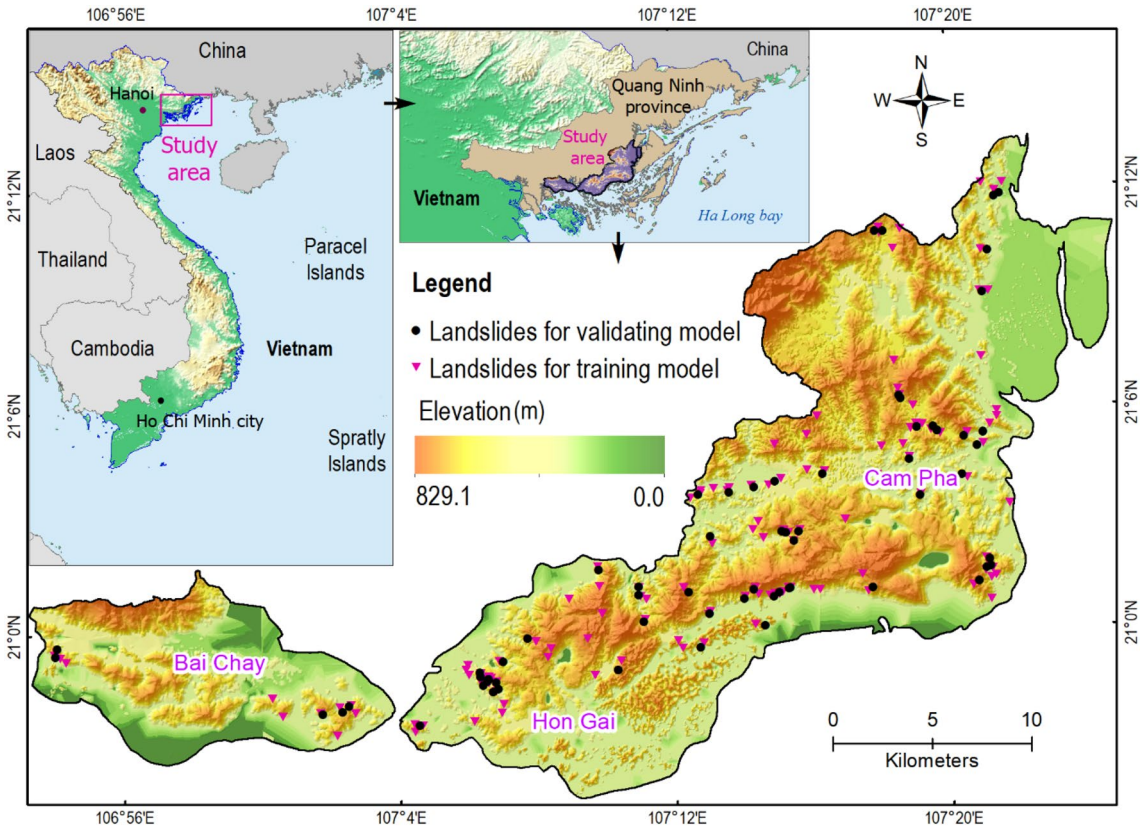

Fig. 3 The study area in Vietnam

Influencing factors

Determination of influencing factors is essential for prone landslide susceptibility mapping. In these study areas, these shallow landslides were soil and rock mixed soil failures and triggered by heavy rainfalls; therefore, land-use, distance to road, distance to river, soil type, distance to fault, and lithology, and the elevation-related factors were considered (Bui et al. 2017a; Dang et al. 2020; Nhu et al. 2020). In this analysis, rainfall data were difficult to obtain due to a very low density of monitoring stations; therefore, rainfall patterns were not used.

The digital terrain model (DTM) for the study area was constructed in ArcGIS 10.6 using the available national topographic maps 1:50,000 scale provided by the Ministry of Natural Resouces and Environment (Nhu et al. 2020). Then, these DTM-related factors were prepared.

Slope degree: Slope is the most important predisposing factor in triggering shallow landslides because it directly controls the balance of motor and resistant forces acting on the soil above the potential rupture surface(Duncan et al. 2014; Griffiths and Fenton 2000). In this research, the slope degree map generated from DTM in the study has a parameter range from 0 to 76.73 degrees (Fig. 4a).

Aspect: The slope aspect describes the orientation of the slope and has a primary effect on precipitation, wind, and sunlight offering (He et al. 2019). As a result, this factor has often been used by researchers in landslide sensitivity analysis. In this study, the slope aspect map of DTM was prepared and had nine parameters (Fig. 4b): flat, four cardinal directions: north, south, east, west, and four intercardinal directions: northeast (NE), southeast (SE), southwest (SW), and northwest (NW).

Curvature: the curvature index is essential in landslide modeling because of its direct association with the hydrological dynamics of the slopes (Beguería 2006; Ozdemir and Altural 2013). Negative numbers indicate concave surfaces or flow concentrator surfaces, whereas positive numbers indicate convex surfaces or flow-dispersing surfaces (Krebs et al. 2015). Numbers close to zero indicate a flat surface. In this study, the curvature map shown in Fig. 4c was derived using the spatial analysis tool in ArcGIS.

Topographic Wetness Index (TWI): TWI is another essential factor in predicting susceptibility to landslides and shows soil conditions, geography, and volume of runoff (He et al. 2019). Its values range from 4.6 to 24.1 (Fig. 4d). In other words, TWI is a topographic factor in

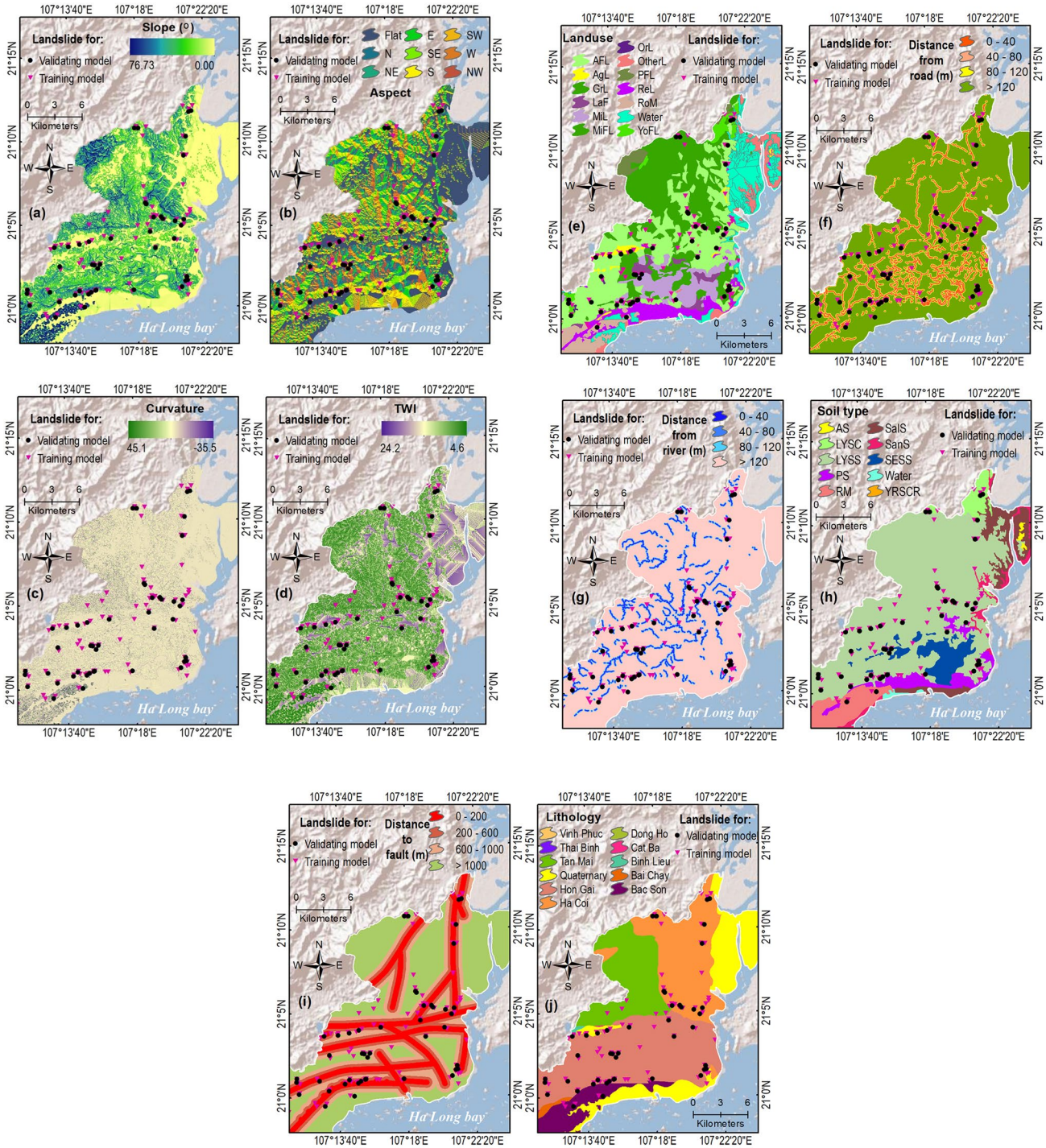

Fig. 4 Landslide related factor: **a** slope, **b** aspect, **c** curvature, **d** topographic wetness index (TWI), **e** land-use (*AFL* Acacia land, *AgL* Agricultural land, *GrL* Grassland, *LaF* landfills, *MiL* Mineral land, *MiFL* Mixed forest land, *OrL* Orchard land; *OtherL* Other lands, *PFL* Poor forest land, *ReL* Residential land, *RoM* Rocky mountain, *YoL* Young forest land), **f** distance to road, **g** distance to river, **h** soil type (*AS* Alluvial soil, *LYSC* Light yellowish soil on claystone and metaphormic rock, *LYSS* Light yellowish soil on sandstone, *PS* Peaty soil, *RM* Rocky mountain, *SalS* Saline soil, *SanS* Sandy soil, *SESS* Strongly eroded skeletal soil, and *YRSCR* Yellowish red soil cultivated with floating rice), (i) distance to fault, and (j) lithology

the runoff model and is defined by the following equation (Moore and Grayson 1991):

$$TWI = \ln\left(\frac{a}{tgb}\right), \quad (2)$$

where α is the cumulative upslope area surface that is bored through a point and $b = \tan\beta$ is the slope angle at the point.

Land use (LU): Land use in slope instability is of particular importance and is widely regarded as susceptible to landslides (Zhao et al. 2015). Explanation Land use types have been generalized to different categories to assess their potential for landslide in Fig. 4e.

Distance to road (DTRo): This factor was considered because our fieldworks show that road sections undercutting slopes influence landslide in this study area. Therefore, first, road-cut sections that undercut slopes larger than 10° were extracted and then used to compute the distance to road map (Fig. 4f).

Distance to river (DTRi): It is recognized that river and stream sections undercutting slopes may influence slope failures indirectly (Bui et al. 2012). Therefore, in this work, river section cutting slopes larger than 10° were used to construct the distance to river map. The buffer map was created using ArcGIS software and then divided into different groups at 40-m intervals: 0–40, 40–80, 80–120, > 120 m (Fig. 4g).

Soil type (ST): Soil type directly affects soil drainage, erosion, and destruction; therefore, it influences shallow landslides (Frattini et al. 2004). The soil feature map in this study area is represented by ten categories (Fig. 4h).

Distance to fault (DTF): Fault distances were divided into different groups: 0–200, 200–600, 600–1000, > 1000 m (Fig. 4i). Fault levels lightly become sliding surfaces because the stress on the rock around a fault is unstable. Landslides often happen during surface rupture (Yalcin et al. 2011).

Lithology type (LT): the lithological map contains 11 separate compositions as shown in Fig. 2j. Undoubtedly, the type of lithology has a profound effect on landslides because it directly affects some essential characteristics such as shear strength and permeability of rocks and soils and, thus, their resistance to weathering/erosion process (Varnes 1984; Watakabe and Matsushi 2019).

Proposed ICA-NeuralNet for predicting shallow landslide susceptible areas

Feature relevancy investigation

In machine learning, the quality of the model depends heavily on the quantity and quality of the input data (Emary et al. 2016). Therefore, it is necessary to evaluate the predictability of factors used in this study preliminarily. The objective of this evaluation is to identify irrelevant variables that can be cast out without incurring a significant loss of information (James et al. 2013). In this research, the wrapper algorithm (Kohavi and John1997) with fivefold cross-validation was employed to compute the merit value, which was used to measure the importance of the landslide influencing factors.

Using the training dataset, a subset of the influencing factors was searched, and then, a landslide susceptibility model was built using the random forests algorithm. In the next step, the performance of the model was assessed using statistical metrics. This is a repeated process, and all the possible combinations of the landslide influencing factors were assessed, and finally, the role of each factor was ranked.

Because the use of statistical metrics may influence the final ranking of the influencing factors, therefore, in this work, four metrics were considered, including classification accuracy (CAR), F-score, the area under the curve (AUC), and mean absolute error (MAE) (Hand 2009). The purpose of using the four metrics is to see if any bias in the conclusion of the feature relevancy.

Designing and configuring ICA- NeuralNet model for shallow landslide

The hybrid model (refer to Fig. 5) was established based on the ICA and the neural computing approach of the NeuralNet. The model structure is arranged in a multi-layered configuration containing an input layer, having one or more intermediate layers (referred to as hidden layers), and an output layer. In this network, the information of different factors is interconnected to exchange information in a unidirectional way starting from the input layer through hidden layers to the output layer (Rodriguez-Galiano et al. 2015).

The ICA is employed to enhance the learning capability of the modeling approach (refer to Fig. 6). This algorithm computes the error between the outputted value and real target value and feeds it back to the model in order to adjust the parameters. This repetitive process is performed until the error is sufficiently small, or the maximum number of epochs is reached. Based on this iterative scheme, the values can be adjusted to yield the desired output that is close enough to the real target information (Çelik and Başarır 2017). The ICA-based optimization process terminates when the maximum number of searching iteration = 100 is achieved.

As shown in the previous section, a NeuralNet model used for shallow landslide mapping is fully determined by IW and OW's two-weight matrices. The first matrix is of the size N_R x $N_I + 1$, where N_R and N_I denote the number of neurons in the hidden layer and the number of input features, respectively. It is worth to mention that the number of columns of IW is $N_I + 1$, representing a vector of bias. Since there are ten landslide influencing factors,

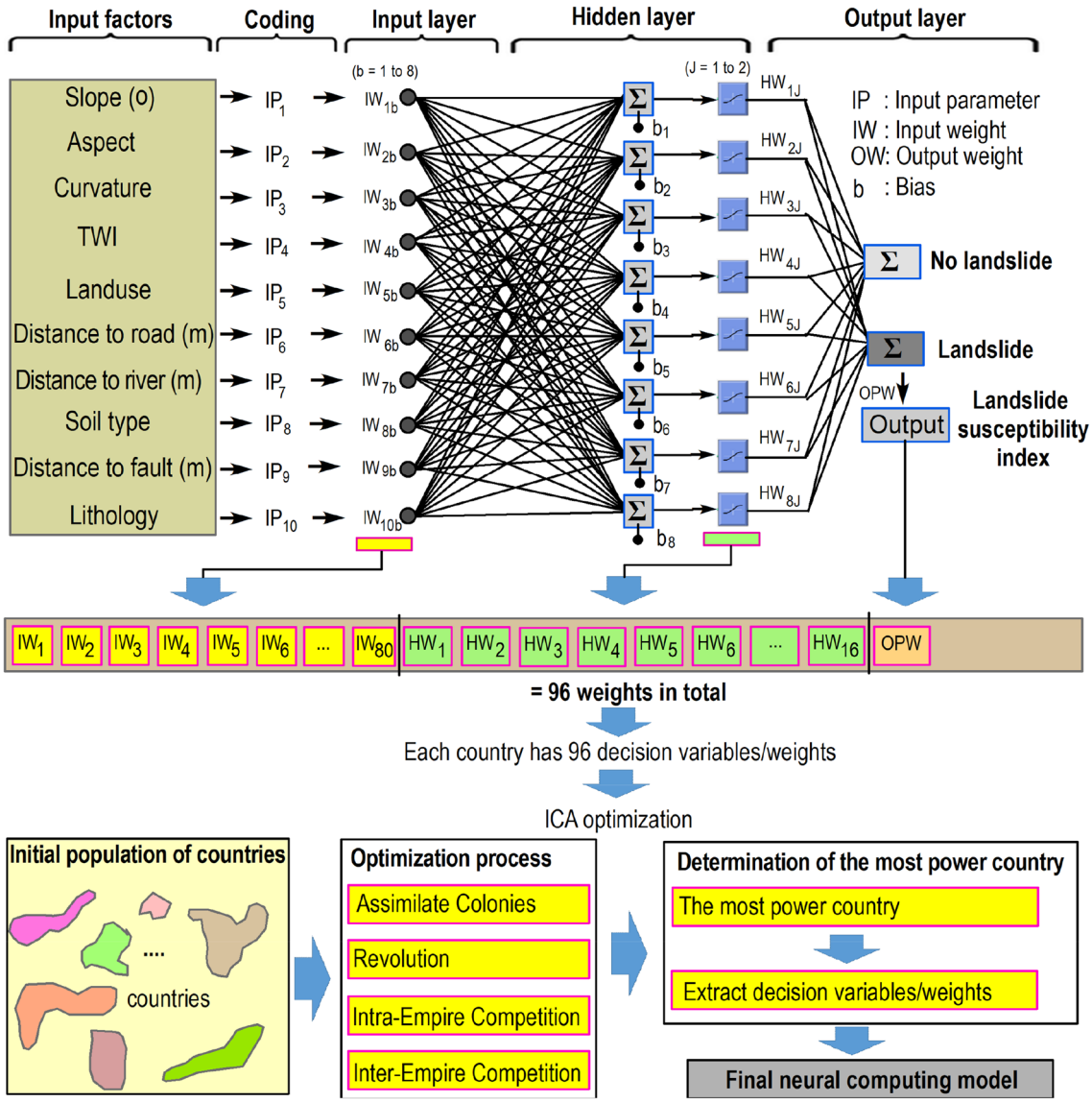

Fig. 5 The proposed ICA-NeuralNet structure in this research

$N_I = 10$. Moreover, the number of neurons in the hidden layer should be determined appropriately via several trial-and-error runs. The second weight matrix OW is of the size N_O x $N_R + 1$. Because the ICA metaheuristic necessitates that a candidate solution is expressed in the form of a vector, the two matrices IW and OW must be vectorized. Thus, the total number of searched variables is N_R x $N_I + N_O$ x $N_R + 2$. Furthermore, to identify the most desired NN model structure, the Root Mean Square Error (RMSE) is used as the objective function to be minimized by the ICA. The RMSE index is computed as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_i - t_i)^2}{N}}, \tag{3}$$

where y_i and t_i are the predicted, and the actual class outputs (either "landslide" or "non-landslide"); N denotes the number of data samples.

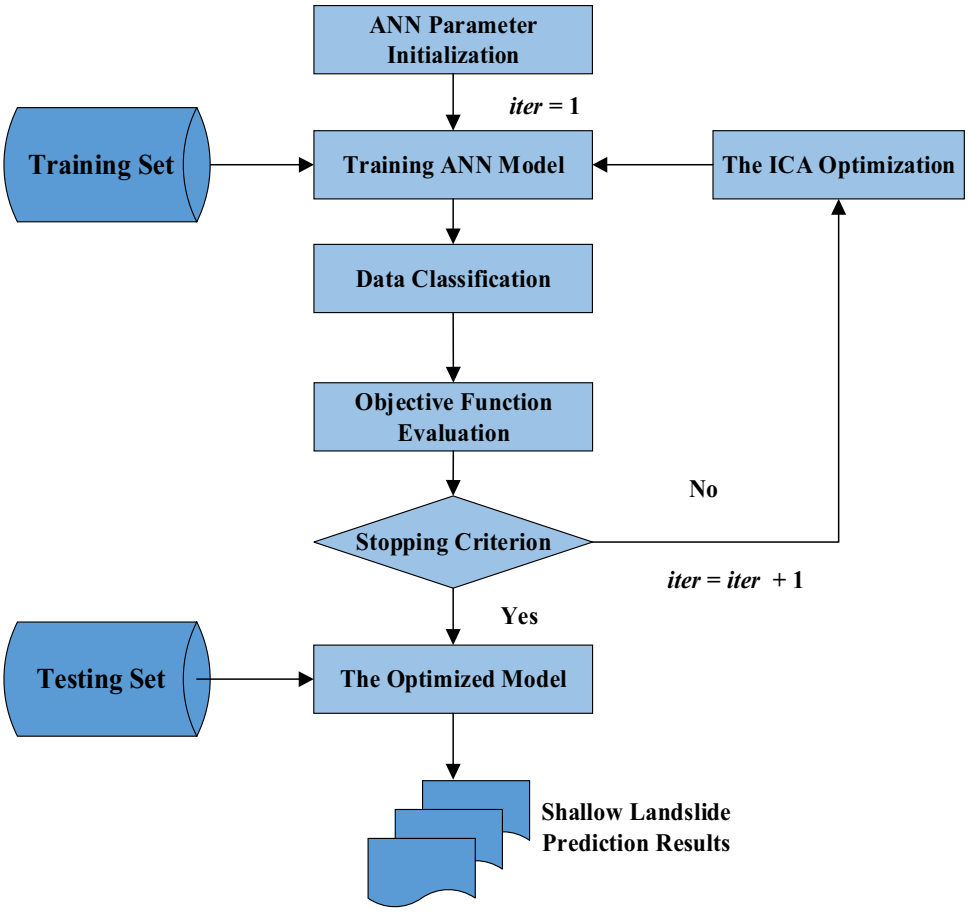

Fig. 6 The ICA-NET optimization process

The ICA-NeuralNet model performance measurement

To assess the ICA-NeuralNet model used for spatial prediction of shallow landslides, the indices of true positive index (TPI), true-negative index (TNI), false-positive index (FPI), and false-negative index (FNI) are first obtained and computed directly by inspecting the classification outcomes. Accordingly, the statistical indices of positive predictive value (PPV), negative predictive value (NPV), classification accuracy (CAR), sensitivity, and specificity can be computed based on the TP mentioned above, TN, FP, and FN as follows (Bui et al. 2016b; Liu et al. 2005):

$$\text{PPV} = \frac{TPI}{TPI + FPI}, \tag{4}$$

$$\text{NPV} = \frac{TNI}{TNI + FNI}, \tag{5}$$

$$\text{CAR} = \frac{TPI + TNI}{TPI + TNI + FPI + FNI} \times 100\%, \tag{6}$$

$$\text{Sensitivity} = \frac{TPI}{TPI + FNI}, \tag{7}$$

$$\text{Specificity} = \frac{TNI}{TNI + FPI}. \tag{8}$$

Besides, one of the most popular methods which are usually used to assess the machine learning models is the Receiver operating characteristic (ROC) curve (van Erkel and Pattynama 1998) and the Kappa index (K) (McHugh 2012). Notably, the area under the ROC Curve (AUC) is usually computed and graphically describes the capability of landslide prediction (Bui et al. 2017b; Hoang and Tien Bui 2016; Lucà et al. 2011). Accordingly, the performance of the model is excellent (AUC belongs to 0.9–1), good (AUC belongs to 0.8–0.9), fair (AUC belongs to 0.7–0.8), and poor (AUC is less than 0.7) (Cantor and Kattan 2000).

Table 1 The role of landslide influencing factors in this study area using the random forest based wrapper subset assessment

No	Influencing factor	Merit value				Ranking
		CAR	F-score	AUC	MAE	
1	Slope (o)	0.2357	0.2274	0.2529	0.1629	1
2	Aspect	0.1661	0.1572	0.1973	0.0985	2
3	TWI	0.1605	0.1597	0.1995	0.0873	3
4	Lithology	0.1618	0.1585	0.1909	0.0810	4
5	Soil	0.1185	0.0897	0.1177	0.0591	5
6	Landuse	0.1144	0.0785	0.1294	0.0453	6
7	Distance to fault (m)	0.1293	0.1338	0.1477	0.0407	7
8	Curvature	0.1339	0.1359	0.1372	0.0392	8
9	Distance to road (m)	0.0856	0.0765	0.0843	0.0184	9
10	Distanc to river (m)	0.0187	0.0759	0.0184	0.0018	10

Table 2 Performance of the ICA-NeuralNet model on the training dataset

Metrics	RF	J48 DT	CART	LMT	ICA-NeuralNet
TPI	1285	1236	1241	1219	1221
TNI	1141	1124	1095	1144	1023
FPI	76	125	120	142	140
FNI	220	237	266	217	338
PPV (%)	94.4	90.8	91.2	89.6	89.7
NPV (%)	83.8	82.6	80.5	84.1	75.2
Sensitivity (%)	85.4	83.9	82.3	84.9	78.3
Specificity (%)	93.8	90.0	90.1	89.0	88.0
CAR (%)	89.1	86.7	85.8	86.8	82.4
K	0.783	0.734	0.716	0.736	0.649
AUC	0.964	0.914	0.891	0.908	0.895

Table 3 Prediction performance of the model in the validating dataset

Metrics	RF	J48 DT	CART	LMT	ICA-NeuralNet
TPI	278	274	272	270	400
TNI	433	426	418	430	388
FPI	226	230	232	234	104
FNI	71	78	86	74	116
PPV (%)	55.2	54.4	54.0	53.6	79.4
NPV (%)	85.9	84.5	82.9	85.3	77.0
Sensitivity (%)	79.7	77.8	76.0	78.5	77.5
Specificity (%)	65.7	64.9	64.3	64.8	78.9
CAR (%)	70.5	69.4	68.5	69.4	78.2
K	0.411	0.389	0.369	0.389	0.563
AUC	0.860	0.740	0.814	0.778	0.847

Result and discussion

Preliminary analysis on feature importance

As mentioned earlier, the two approaches of mutual information and ReliefF are employed in this study to preliminarily investigate the relevancy of the ten shallow landslide influencing factors. The investigation outcomes are reported in Table 1. It can be seen from this table that the slope factor has obtained the highest merit value in all cases of the four different metrics used (CAR, F-score, AUC, and MAE). This is plausible because the slope is widely accepted as the most critical factor causing failures (Bollati et al. 2012; Günther et al. 2013; Hong et al. 2007). In contrast, distance to road and distance to river have the lowest contribution to the landslide. The ranking of the other seven factors is slightly different with the four statistical metrics (Table 1). Herein, three factors, aspect, TWI, and lithology, have higher contributions than soil, land-use, distance to fault, and curvature. All factors have a certain merit value; therefor,e they were employed all for the training phase in the next step.

Training ICA-NeuralNet model and performance assessment

The ICA-NeuralNet model was constructed by the training dataset created from ten predisposing factors. The training was accomplished via a tenfold cross-validation process. The training result is reported in Table 2. It can be seen clearly that the ICA-NeuralNet model can provide a good fit to the training dataset with CAR = 82.4%, K = 0.649, and AUC = 0.895). After the ICA-NeuralNet model training phase is accomplished, the validation dataset is used to test the model predictive capability. The validation result is summarized in Table 3 with CAR = 78.2%, sensitivity = 77.5%, specificity = 78.9%, K = 0.563, and AUC = 0.847).

In addition, to better demonstrate the predictive capability of the ICA-NeuralNet, the verification of the model was taken further by two indexes, the success-rate curve and the prediction-rate method (Bui et al. 2016a; Chung and Fabbri 2003). The result of the success-rate curve is illustrated in

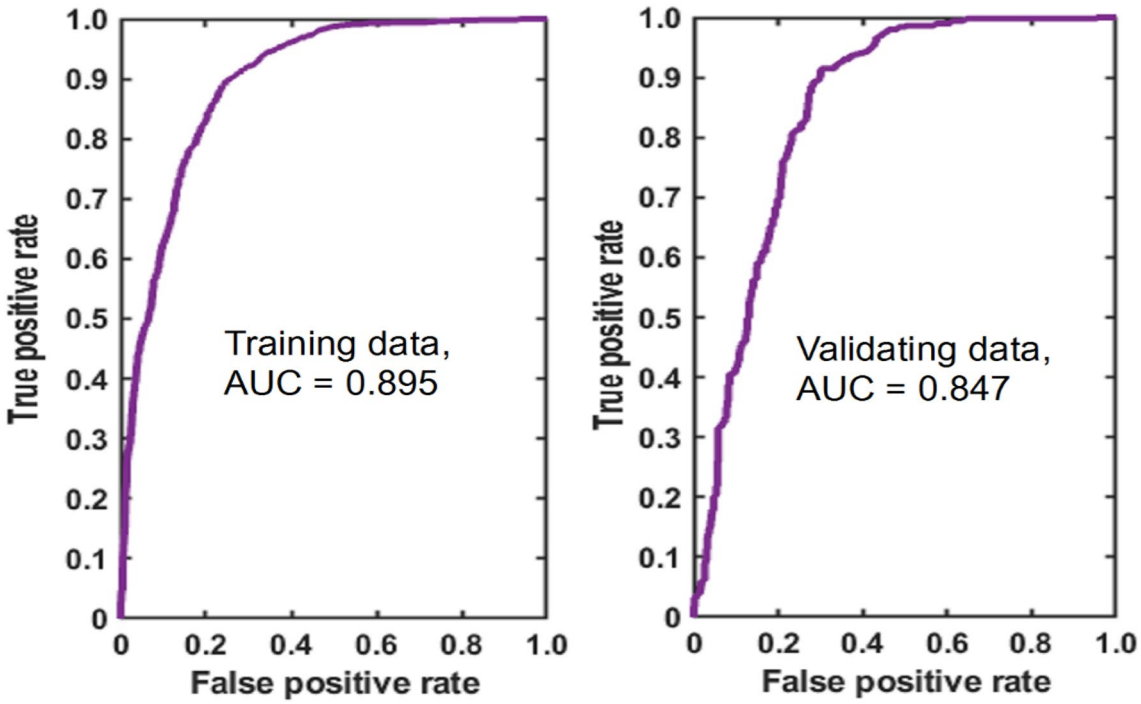

Fig. 7 ROC curve and AUC of the ICA-NeuralNet model

Fig. 7, which was constructed by comparing the landslide susceptibility indices with landslide pixels in the training dataset and the validation dataset. The AUC values obtained from the ICA-NeuralNet are 0.895 for the training phase and 0.847 for the validation phase. These outcomes confirm that the model has high-prediction capability.

Comparison of the ICA-NeuralNet model with the benchmark

As mentioned earlier, the ICA-NeuralNet model was proposed in this study by combining the ICA metaheuristic and neural computing model for landslide modeling. To demonstrate the advantage of the newly developed model, the ICA-NeuralNet is compared with the benchmark models, including the Random Forest (RF), J48 DT, Classification and regression trees (CART), and logistic model tree (LMT). It is because those methods have been used successfully in establishing landslide susceptibility maps in previous works (Bui et al. 2016a, b; Chen et al. 2017; Felicísimo et al. 2013; Hong et al. 2018; Youssef et al. 2016).

With a focus on validating performances, the landslide prediction results of the proposed ICA-NeuralNet model (CAR = 78.2%, $K = 0.563$, and AUC = 0.847) are better than those of the benchmark methods. The RF is the second-best approach with CAR = 70.5%, $K = 0.411$, and AUC = 0.860) followed by J48 DT, CAR, and LMT. More importantly, the ICA-NeuralNet has a good balance in processing both landslide and non-landslide pixels, where the sensitivity and the specificity are 77.5% and 78.9%, respectively. In contrast, the other models (RF, J48 DT, CART, and LMT) have some bias in processing landslide pixels, where FPI is significantly higher than the proposed ICA-NeuralNet model (Table 3). Thus, it could be concluded that the ICA-NeuralNet model is highly effective and suited for spatial modeling of the shallow landslide in the study area.

Shallow landslide susceptibility map

When the final ICA-NeuralNet model is established, the model is then used to calculate the sensitivity index for all pixels of the study area. Accordingly, these susceptibility indices were converted to the ASCII raster format in ArcGIS using Python. Finally, the landslide susceptibility map was created for the study area. The IC-NET model optimized by the ICA metaheuristic is then used to calculate landslide susceptibility indices for the study area. All of the influencing factors were converted to raster format and then fed to the ICA-NeuralNet model to generate susceptibility indices called landslide probability index. These indexes were classified based on the influence level of the factors to landslide probability occurrence or susceptibility. Accordingly, the landslide susceptibility map (shown in Fig. 8) for the area was cartographically presented by values ranging from 0 to 1.

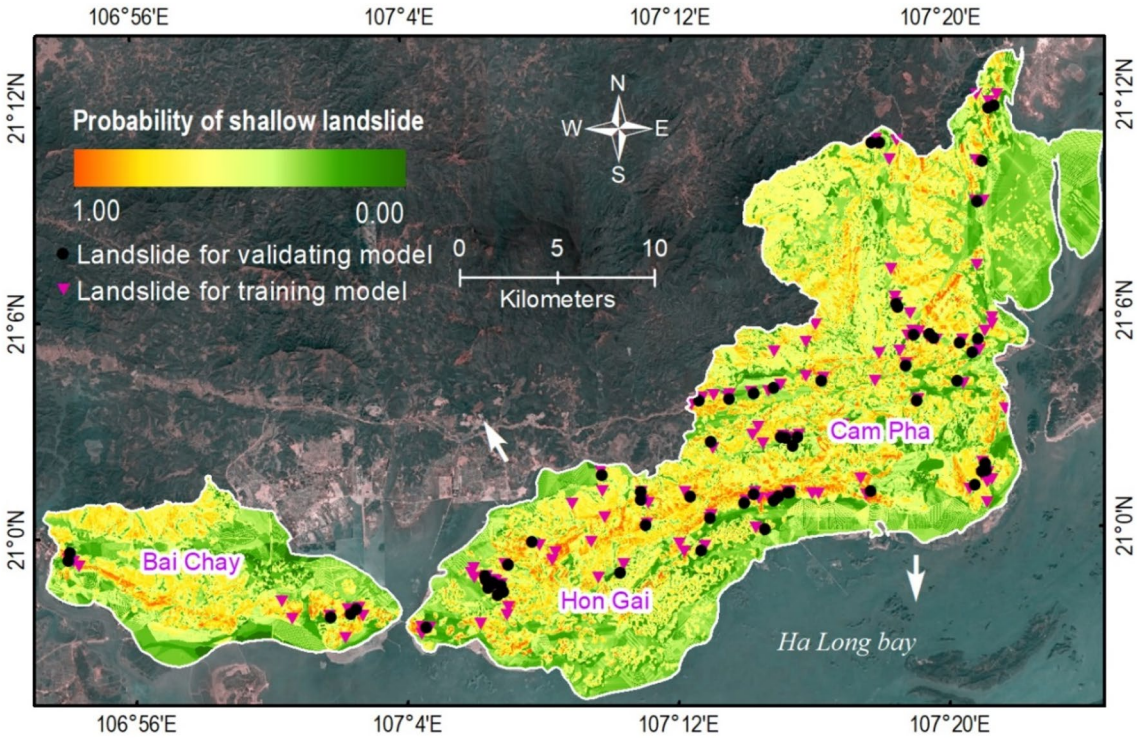

Fig. 8 Shallow landslide susceptibility map generated by the proposed ICA-NeuralNet model

Concluding remarks

Due to recent climate change and land-use modification, landslide occurrences in the study area of Ha Long and Cam Pha cities (northern Vietnam) are escalating in both scales of damage and frequency. Therefore, establishing a shallow landslide susceptibility map is crucial and an urgent need. This study proposed a novel machine learning solution for constructing such a landslide susceptibility map for the study area by hybridizing the metaheuristic method of the ICA and neural computing framework. The hybrid method, named as ICA-NeuralNet, combines the advantages of NeuralNet as a robust nonlinear classifier and of the ICA as a global optimization. The NeuralNet plays a role in building a decision boundary separating the input space into two distinct regions of "landslide" and "non-landslide". Meanwhile, the ICA plays a role in optimizing the NeuralNet model structure with respect to the collected GIS dataset. Since the ICA is a stochastic global optimizer, it can help the training phase of the NeuralNet to avoid being trapped in an optimal local solution.

Training and validating the proposed method were accomplished based on a GIS dataset, including 3730 data samples and ten landslide influencing factors. Experimental results point out that the ICA can help to avoid the risk of premature convergence and attain a high-quality NeuralNet model with CAR more than 78% and AUC = 0.85. Thus, the results of this study have illustrated the effectiveness of using ICA to optimize the neural computing model for predicting shallow landslide susceptibility. Hence, the proposed model has a good potentiality to be used for susceptibility mapping of the shallow landslide in other areas that have the same geo-environment conditions. Finally, this study's results could be used for further research, including land-use planning/management in landslide-prone areas in Quang Ninh province (Vietnam).

Since the ICA-NeuralNet model has demonstrated good predictive accuracy for the study area, the proposed approach can be potentially applied for shallow landslide assessment in other regions. Nevertheless, since the model prediction performance is highly dependent on the quality and quantity of the GIS data, a decent effort on data collection is required to ensure the successful application of the hybrid metaheuristic-neural computation model. One limitation of the current study is that the hyper-parameters of the neural network model, including the number of neurons in the hidden layer and the activation function, are selected via a trial-and-error process. Therefore, the current work's future extensions may investigate an advanced approach for automatic fine-tune such as hyper-parameters of the neural computing models. The current work has investigated and confirmed the capability of the ICA-optimized neural network. Future works may explore

other neural computing models, such as convolutional neural networks for shallow landslide susceptibility.

Acknowledgement This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant number 105.08-2017.316

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest.

References

Achour Y, Pourghasemi HR (2020) How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? Geosci Front 11:871–883. https://doi.org/10.1016/j.gsf.2019.10.001

Aditian A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. Geomorphology 318:101–111. https://doi.org/10.1016/j.geomorph.2018.06.006

Aghdam IN, Pradhan B, Panahi MJ (2017) Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy inference system (ANFIS) at southern Zagros Mountains in Iran. Environ Earth Sci 76:237

Atashpaz-Gargari E, Lucas C (2007) Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. 2007 IEEE congress on evolutionary computation. IEEE. pp 4661–4667

Bandara A, Hettiarachchi Y, Hettiarachchi K, Munasinghe S, Wijesinghe I, Thayasivam U (2020) A generalized ensemble machine learning approach for landslide susceptibility modeling. In: Sharma N, Chakrabarti A, Balas VE (eds) Data Management, Analytics and Innovation. Springer Singapore, Singapore, pp 71–93

Barraqué B, Moatty A (2019) The French Cat'Nat'system: post-flood recovery and resilience issues. Environ Hazards.1–16

Beguería S (2006) Changes in land cover and shallow landslide activity: a case study in the Spanish Pyrenees. Geomorphology 74:196–206

Benz SA, Blum P (2019) Global detection of rainfall-triggered landslide clusters. Nat Hazards Earth Syst Sci 19:1433–1444

Bishop CM (2011) Pattern Recognition and Machine Learning (Information Science and Statistics) Springer (April 6, 2011). ISBN- 10:0387310738

Bollati I, Della Seta M, Pelfini M, Del Monte M, Fredi P, Palmieri EL (2012) Dendrochronological and geomorphological investigations to assess water erosion and mass wasting processes in the Apennines of Southern Tuscany (Italy). CATENA 90:1–17

Bui DT, Ho T-C, Pradhan B, Pham B-T, Nhu V-H, Revhaug IJEES (2016a) GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost. Bagging MultiBoost Ensemble Frameworks 75:1101

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. Geomorphology 171:12–29

Bui DT, Tuan TA, Hoang N-D, Thanh NQ, Nguyen DB, Van Liem N, Pradhan B (2017a) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14:447–458

Bui DT, Tuan TA, Hoang N-D, Thanh NQ, Nguyen DB, van Liem N, Pradhan BJL (2017) Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14:447–458

Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug IJL (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13:361–378

Cantor SB, Kattan MWJMDM (2000) Determining the area under the ROC curve for a binary diagnostic test. Med Decis Mak 20:468–470

Cao J, Zhang Z, Wang C, Liu J, Zhang L (2019) Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. CATENA 175:63–76. https://doi.org/10.1016/j.catena.2018.12.013

Cascini LJEG (2008) Applicability of landslide susceptibility and hazard zoning at different scales. Eng Geol. 102:164–177

Çelik U, Başarır ÇJAJ (2017) The prediction of precious metal prices via artificial neural network by using RapidMiner. Alphanum J 5:45–54

Chatterjee S, Sarkar S, Hore S, Dey N, Ashour AS, Balas VE (2017) Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. Neural Comput Appl 28:2005–2016. https://doi.org/10.1007/s00521-016-2190-2

Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S, Li S, Jaafari A, Ahmad BB (2019) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. CATENA 172:212–231. https://doi.org/10.1016/j.catena.2018.08.025

Chen W, Peng J, Hong H, Shahabi H, Pradhan B, Liu J, Zhu AX, Pei X, Duan Z (2018) Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Sci Total Environ 626:1121–1135. https://doi.org/10.1016/j.scitotenv.2018.01.124

Chen W, Pourghasemi HR, Zhang S, Wang J (2019) A Comparative Study of Functional Data Analysis and Generalized Linear Model Data-Mining Techniques for Landslide Spatial Modeling. Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, New York, pp 467–484

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma JJC (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena. 151:147–160

Chung C-JF, Fabbri AGJNH (2003) Validation of spatial prediction models for landslide hazard mapping. Natural Hazards 30:451–472

Ciurleo M, Mandaglio MC, Moraci N (2019) Landslide susceptibility assessment by TRIGRS in a frequently affected shallow instability area. Landslides 16:175–188. https://doi.org/10.1007/s10346-018-1072-3

Dang V-H, Hoang N-D, Nguyen L-M-D, Bui DT, Samui P (2020) A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility. Forests 11:118

Ding Q, Chen W, Hong H (2017) Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto Int 32:619–639. https://doi.org/10.1080/10106049.2016.1165294

Ding Z, Nguyen H, Bui X-N, Zhou J, Moayedi H (2019) Computational intelligence model for estimating intensity of

blast-induced ground vibration in a mine based on imperialist competitive and extreme gradient boosting algorithms. Nat Resour Res. https://doi.org/10.1007/s11053-019-09548-8
Duncan JM, Wright SG, Brandon TL (2014) Soil strength and slope stability. Wiley
Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381. https://doi.org/10.1016/j.neucom.2015.06.083
Felicísimo ÁM, Cuartero A, Remondo J, Quirós EJL (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10:175–189
Fenart P, Cat N, Drogue C, Van Canh D, Pistre S (1999) Influence of tectonics and neotectonics on the morphogenesis of the peak karst of Halong Bay. Vietnam Geodinamica Acta 12:193–200
Frattini P, Crosta GB, Fusi N, Dal Negro P (2004) Shallow landslides in pyroclastic soils: a distributed modelling approach for hazard assessment. Eng Geol 73:277–295
Fressard M, Thiery Y, Maquaire O (2014) Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d'Auge plateau hillslopes. Normandy, France
Froude MJ, Petley DJNH, Sciences ES (2018) Global fatal landslide occurrence from 2004 to 2016. Nat Hazards Earth Syst Sci 18:2161–2181
Gerist S, Maheri MR (2019) Structural damage detection using imperialist competitive algorithm and damage function. Appl Soft Comput 77:1–23. https://doi.org/10.1016/j.asoc.2018.12.032
Ghaleini EN, Koopialipoor M, Momenzadeh M, Sarafraz ME, Mohamad ET, Gordan B (2019) A combination of artificial bee colony and neural network for approximating the safety factor of retaining walls. Eng Comput 35:647–658. https://doi.org/10.1007/s00366-018-0625-3
Grahn T, Jaldell HJL (2017) Assessment of data availability for the development of landslide fatality curves. Landslides 14:1113–1126
Griffiths D, Fenton GA (2000) Influence of soil strength spatial variability on the stability of an undrained clay slope by finite elements. Slope Stab 2000:184–193
Guru B, Veerappan R, Sangma F, Bera SJSIR (2017) Comparison of probabilistic and expert-based models in landslide susceptibility zonation mapping in part of Nilgiri District, Tamil Nadu. India 25:757–768
Günther A, Reichenbach P, Malet J-P, Van Den Eeckhaut M, Hervás J, Dashwood C, Guzzetti F (2013) Tier-based approaches for landslide susceptibility assessment in Europe. Landslides 10:529–546
Göçken M, Özçalıcı M, Boru A, Dosdoğru AT (2016) Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. Expert Syst Appl 44:320–331. https://doi.org/10.1016/j.eswa.2015.09.029
Ha ND, Sayama T, Sassa K, Takara K, Uzuoka R, Dang K, Van Pham T (2020) A coupled hydrological-geotechnical framework for forecasting shallow landslide hazard—a case study in Halong City. Vietnam. Landslides. 1:1–16
Hand DJ (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve. Mach Learn 77:103–123
Haque U, Da Silva PF, Devoli G, Pilz J, Zhao B, Khaloua A, Wilopo W, Andersen P, Lu P, Lee J (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). Sci Total Environ 682:673–684
Harmouzi H, Nefeslioglu HA, Rouai M, Sezer EA, Dekayir A, Gokceoglu C (2019) Landslide susceptibility mapping of the Mediterranean coastal zone of Morocco between Oued Laou and El Jebha using artificial neural networks (ANN). Arab J Geosci 12:696. https://doi.org/10.1007/s12517-019-4892-0
He Q, Shahabi H, Shirzadi A, Li S, Chen W, Wang N, Chai H, Bian H, Ma J, Chen YJSotTE, (2019) Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. Sci Total Environ 663:1–15
Hegazy T, Fazio P, Moselhi O (1994) Developing practical neural network applications using back-propagation. Comput Aided Civ Inf 9:145–159. https://doi.org/10.1111/j.1467-8667.1994.tb00369.x
Hoang N-D, Tien Bui D (2017) Chapter 18 - Slope stability evaluation using radial basis function neural network, least squares support vector machines, and extreme learning machine. In: Handbook of Neural Computation. Academic Press. pp 333–344
Hoang N-D, Tien Bui DJ (2016) A novel relevance vector machine classifier with cuckoo search optimization for spatial prediction of landslides. J Comput Civil Eng 30:04016001
Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu A-X, Chen W, Ahmad BBJC (2018) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena 163:399–413
Hong Y, Adler R, Huffman G (2007) Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat Hazards 43:245–256
Hosseini S, Al Khaled A (2014) A survey on the Imperialist Competitive Algorithm metaheuristic: Implementation in engineering domain and directions for future research. Appl Soft Comput 24:1078–1094. https://doi.org/10.1016/j.asoc.2014.08.024
Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. CATENA 165:520–529. https://doi.org/10.1016/j.catena.2018.03.003
James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning. Springer, New York, NY
Keellings D, Hernández Ayala JJ (2019) Extreme rainfall associated with Hurricane Maria over Puerto Rico and its connections to climate variability and change. Geophys Res Lett 46:2964–2973
Kohavi R, John GH (1997) Wrappers for feature subset selection. Artif Intell 97:273–324
Kornejady A, Pourghasemi HR, Afzali SF (2019) Presentation of RFFR New Ensemble Model for Landslide Susceptibility Assessment in Iran. In: Pradhan SP, Vishal V, Singh TN (eds) Landslides: Theory, Practice and Modelling. Springer International Publishing, Cham, pp 123–143
Kowalski PA, Łukasik S (2016) Training Neural Networks with Krill Herd Algorithm. Neural Process Lett 44:5–17. https://doi.org/10.1007/s11063-015-9463-0
Krebs P, Stocker M, Pezzatti GB, Conedera M (2015) An alternative approach to transverse and profile terrain curvature. Int J Geogr Inf Sci 29:643–666
Kuriakose SL, Van Beek L, Westen Van, CJEsp, landforms, (2009) Parameterizing a physically based shallow landslide model in a data poor region. Earth Surface Process Landforms. 34:867–881
Kutlug Sahin E, Colkesen I (2019) Performance analysis of advanced decision tree-based ensemble learning algorithms for landslide susceptibility mapping. Geocarto Int. https://doi.org/10.1080/10106049.2019.1641560
Le HV, Bui QT, Tien Bui D, Tran HH, Hoang ND (2018) A hybrid intelligence system based on relevance vector machines and imperialist competitive optimization for modelling forest fire danger using GIS. J Environ Inform. https://doi.org/10.3808/jei.201800404
Li J, Luo Q, Liao L, Zhou Y (2018) Using Spotted Hyena Optimizer for Training Feedforward Neural Networks. Springer, Cham, pp 828–833
Li L, Lan H, Guo C, Zhang Y, Li Q, Wu Y (2017) A modified frequency ratio method for landslide susceptibility assessment. Landslides 14:727–741. https://doi.org/10.1007/s10346-016-0771-x

Liu C, Berry PM, Dawson TP, Pearson RGJE (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography. 28:385–393

Liu M, Yang L, Smith J, Vecchi G (2019) Response of extreme rainfall for landfalling tropical cyclones undergoing extratropical transition to projected climate change: Hurricane Irene (2011). Earth's Fut 8:e2019EF001360

Loi DH, Quang LH, Sassa K, Takara K, Dang K, Thanh NK, Van Tien P (2017) The July 28 2015 rapid landslide at Ha Long City, Quang Ninh. Vietnam Landslides 14:1207–1215

Lucà F, Conforti M, Robustelli GJG (2011) Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. Geomorphology 134:297–308

Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA (2019) Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J Photogramm Remote Sens 152:166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22:276–282

Merghadi A, Abderrahmane B, Tien Bui D (2018) Landslide Susceptibility Assessment at Mila Basin (Algeria): a comparative assessment of prediction capability of advanced machine learning methods. ISPRS Int J Geo-Inform 7:268

Mikaeil R, Haghshenas SS, Haghshenas SS, Ataei M (2018) Performance prediction of circular saw machine using imperialist competitive algorithm and fuzzy clustering technique. Neural Comput Appl 29:283–292. https://doi.org/10.1007/s00521-016-2557-4

Mirjalili SZ, Saremi S, Mirjalili SM (2015) Designing evolutionary feedforward neural networks using social spider optimization algorithm. Neural Comput Appl 26:1919–1928. https://doi.org/10.1007/s00521-015-1847-6

Moayedi H, Tien Bui D, Anastasios D, Kalantar B (2019) Spotted hyena optimizer and ant lion optimization in predicting the shear strength of soil. Appl Sci 9:4738

Mohan A, Singh AK, Kumar B (2020) Dwivedi R Review on remote sensing methods for landslide detection using machine and deep learning. Trans Emerg Telecommun Technol. https://doi.org/10.1002/ett.3998

Moore ID, Grayson RB (1991) Terrain-based catchment partitioning and runoff prediction using vector elevation data. Water Resourc Res. 27:1177–1191

Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng Geol 97:171–191

Nguyen LC, Van Tien P, Do T-N (2020) Deep-seated rainfall-induced landslides on a new expressway: a case study in Vietnam. Landslides 17:395–407

Nhu V-H, Hoang N-D, Nguyen H, Ngo PTT, Bui TT, Hoa PV, Samui P, Bui DT (2020) Effectiveness assessment of keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area. CATENA 188:104458

Ojha VK, Abraham A, Snášel V (2017) Metaheuristic design of feedforward neural networks: a review of two decades of research. Eng Appl Artif Intell 60:97–116. https://doi.org/10.1016/j.engappai.2017.01.013

Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. J Asian Earth Sci 64:180–197

Pham BT, Bui DT, Prakash I, Dholakia MJC (2017) Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena 149:52–63

Pham BT, Bui DT, Prakash IJG, Engineering G (2017) Landslide susceptibility assessment using bagging ensemble based alternating decision trees, logistic regression and J48 decision trees methods: a comparative study. Geotechn Geol Eng 35:2597–2611

Pham VC (2018) Quang Ninh Statistical yearbook, Statistical Publishing House, Quang Ninh Statistics Office

Pham BT, Shirzadi A, Tien Bui D, Prakash I, Dholakia MB (2018) A hybrid machine learning ensemble approach based on a Radial Basis Function neural network and Rotation Forest for landslide susceptibility modeling: A case study in the Himalayan area, India. Int J Sedim Res 33:157–170. https://doi.org/10.1016/j.ijsrc.2017.09.008

Pham BT, Son LH, Hoang T-A, Nguyen D-M, Tien Bui D (2018) Prediction of shear strength of soft soil using machine learning methods. CATENA 166:181–191. https://doi.org/10.1016/j.catena.2018.04.004

Pourghasemi HR, Rahmati OJC (2018) Prediction of the landslide susceptibility: which algorithm, which precision? Catena 162:177–192

Pourghasemi HR, Teimoori Yansari Z, Panagos P, Pradhan B (2018) Analysis and evaluation of landslide susceptibility: a review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). Arab J Geosci 11:193. https://doi.org/10.1007/s12517-018-3531-5

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51:350–365. https://doi.org/10.1016/j.cageo.2012.08.023

Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ Model Softw 25:747–759. https://doi.org/10.1016/j.envsoft.2009.10.016

Quyen H (2015) Rainfall and floods causing damage to Quang Ninh nearly 2,200 billion VND. http://thoibaotaichinhvietnam.vn/pages/thoi-su/2015-08-03/mua-lu-gay-thiet-hai-cho-quang-ninh-uoc-gan-2200-ty-dong-23226.aspx. Accessed 13 Mar 2021

Razavizadeh S, Solaimani K, Massironi M, Kavian A (2017) Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: a case study in northern Iran. Environ Earth Sci 76:499. https://doi.org/10.1007/s12665-017-6839-7

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth Sci Rev 180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas MJOGR (2015) Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev 71:804–818

Rossi M, Guzzetti F, Salvati P, Donnini M, Napolitano E, Bianchi CJESR (2019) A predictive model of societal landslide risk in Italy

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536. https://doi.org/10.1038/323533a0

Sachdeva S, Bhatia T, Verma AK (2020) A novel voting ensemble model for spatial prediction of landslides using GIS. Int J Remote Sens 41:929–952. https://doi.org/10.1080/01431161.2019.1654141

Sarhadi A, Soulis ED (2017) Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate. Geophys Res Lett 44:2454–2463

Sezer EA, Pradhan B, Gokceoglu C (2011) Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang valley, Malaysia. Expert Syst Appl 38:8208–8219. https://doi.org/10.1016/j.eswa.2010.12.167

Stäubli A, Nussbaumer SU, Allen SK, Huggel C, Arguello M, Costa F, Hergarten C, Martínez R, Soto J, Vargas R (2018) Analysis of weather-and climate-related disasters in mountain regions using different disaster databases. Climate change, extreme events and disaster risk reduction. Springer. pp 17–41

Sundermeyer M, Schlüter R, Ney H (2012) LSTM neural networks for language modeling. Thirteenth annual conference of the international speech communication association

Tashayo B, Behzadafshar K, Soltani Tehrani M, Afkhami Banayem H, Hashemi MH, Taghavi Nezhad SS (2019) Feasibility of imperialist competitive algorithm to predict the surface settlement induced by tunneling. Eng Comput 35:917–923. https://doi.org/10.1007/s00366-018-0641-3

Thanh TD (2012) The Halong Bay Geological Wonde. Vietnam J Earth Sci 34:162–167

Tien Bui D, Hoang N-D, Nguyen H, Tran X-L (2019) Spatial prediction of shallow landslide using Bat algorithm optimized machine learning approach: a case study in Lang Son Province. Vietnam Adv Eng Inform 42:100978. https://doi.org/10.1016/j.aei.2019.100978

Tien Bui D, Shahabi H, Shirzadi A, Chapi K, Hoang N-D, Pham BT, Bui Q-T, Tran C-T, Panahi M, Bin Ahmad B, Saro L (2018) A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. Remote Sensing 10:1538

Truong XL, Mitamura M, Kono Y, Raghavan V, Yonezawa G, Truong XQ, Do TH, Tien Bui D, Lee S (2018) Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. Appl Sci 8:1046

Tsangaratos P, Ilia I, Hong H, Chen W, Xu C (2017) Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China. Landslides 14:1091–1111. https://doi.org/10.1007/s10346-016-0769-4

van Erkel AR, Pattynama PMT (1998) Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology. Eur J Radiol 27:88–94. https://doi.org/10.1016/S0720-048X(97)00157-5

Van NTH, Van Son P, Ninh NH, Tam N, Huyen NT (2017) Landslide inventory mapping in the fourteen Northern provinces of Vietnam: achievements and difficulties. Workshop on World Landslide Forum. Springer. pp 501–510

Varnes DJ (1984) Landslide hazard zonation: a review of principles and practice

Wang H, Zhang L, Yin K, Luo H, Li J (2020) Landslide identification using machine learning. Geosci Front. https://doi.org/10.1016/j.gsf.2020.02.012

Wang L-J, Guo M, Sawada K, Lin J, Zhang J (2016) A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Geosci J 20:117–136. https://doi.org/10.1007/s12303-015-0026-1

Wang L, Zeng Y, Chen T (2015) Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. Expert Syst Appl 42:855–863. https://doi.org/10.1016/j.eswa.2014.08.018

Wang M, Pan G, Liu Y (2019) A novel imperialist competitive algorithm for multithreshold image segmentation. Math Prob Eng 2019:18. https://doi.org/10.1155/2019/5982410

Watakabe T, Matsushi Y (2019) Lithological controls on hydrological processes that trigger shallow landslides: Observations from granite and hornfels hillslopes in Hiroshima, Japan. CATENA 180:55–68

Yalcin A, Reis S, Aydinoglu A, Yomralioglu TJC (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena 85:274–287

Yan G, Liang S, Gui X, Xie Y, Zhao H (2019) Optimizing landslide susceptibility mapping in the Kongtong District, NW China: comparing the subdivision criteria of factors. Geocarto Int 34:1408–1426. https://doi.org/10.1080/10106049.2018.1499816

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MMJL (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13:839–856

Zêzere JL, Pereira S, Melo R, Oliveira SC, Garcia RA (2017) Mapping landslide susceptibility using data-driven methods. Sci Total Environ 1(589):250–67

Zhang J, Hao T, Dong S, Chen X, Cui J, Yang X, Liu C, Li T, Xu Y, Huang S (2015) The structural and tectonic relationships of the major fault systems of the Tan-Lu fault zone, with a focus on the segments within the North China region. J Asian Earth Sci 110:85–100

Zhao C, Chen W, Wang Q, Wu Y, Yang BJAJoG, (2015) A comparative study of statistical index and certainty factor models in landslide susceptibility mapping: a case study for the Shangzhou District, Shaanxi Province, China. Arab J Geosci. 8:9079–9088

Authors and Affiliations

Viet-Ha Nhu[1] · Nhat-Duc Hoang[2,3] · Mahdis Amiri[4] · Tinh Thanh Bui[5] · Phuong Thao T. Ngo[6] · Pham Viet Hoa[7] · Pijush Samui[8] · Long Nguyen Thanh[9] · Tu Pham Quang[10] · Dieu Tien Bui[11]

Nhat-Duc Hoang
hoangnhatduc@duytan.edu.vn

Mahdis Amiri
mahdisamiri94@gmail.com

Tinh Thanh Bui
buitinhhumg@gmail.com

Phuong Thao T. Ngo
ngothiphuongthao@humg.edu.vn

Pham Viet Hoa
pvhoa@hcmig.vast.vn

Pijush Samui
pijush@nitp.ac.in

Long Nguyen Thanh
ntlong1974@yahoo.com

Tu Pham Quang
tupq@tlu.edu.vn

Dieu Tien Bui
dieu.t.bui@usn.no

[1] Department of Geological-Geotechnical Engineering, Hanoi University of Mining and Geology, Hanoi, Vietnam

[2] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

[3] Faculty of Civil Engineering, Duy Tan University, Da Nang 550000, Vietnam

[4] Department of Watershed & Arid Zone Management, Gorgan University of Agricultural Sciences & Natural Resources, Gorgan, Iran

[5] Department of Prospecting and Exploration Geology, Hanoi University of Mining and Geology Hanoi, Hanoi, Vietnam

[6] Faculty of Information Technology, Hanoi University of Mining and Geology Hanoi, Hanoi, Vietnam

[7] Ho Chi Minh City Institute of Resources Geography, Vietnam, Academy of Science and Technology, Mac Dinh Chi 1, Ben Nghe, 1 District, Ho Chi Minh City 700000, Vietnam

[8] Department of Civil Engineering, National Institute of Technology Patna, Patna 800005, India

[9] Department of Economic Geology and Geomatics, Vietnam Institute of Geosciences and Mineral Resources, Hanoi, Vietnam

[10] Geotechnical Engineering Division, Thuyloi University, Hanoi, Vietnam

[11] GIS Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, 3800 Bø, Telemark, Norway