Science of the Total Environment 720 (2020) 137320

Contents lists available at ScienceDirect

Science of the Total Environment

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

Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning

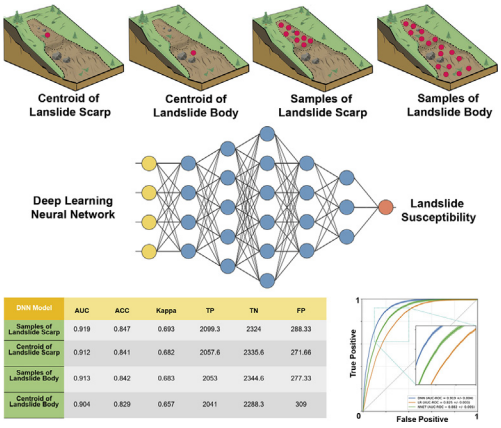
Jie Dou [a,b,*,1], Ali P. Yunus [c,1], Abdelaziz Merghadi [d,1], Ataollah Shirzadi [e], Hoang Nguyen [f], Yawar Hussain [g,h], Ram Avtar [i], Yulong Chen [j,*], Binh Thai Pham [k,*], Hiromitsu Yamagishi [l]

[a] Three Gorges Research Center for Geo-Hazards, Ministry of Education, China University of Geosciences, Wuhan, 430074, China
[b] Department of Civil and Environmental Engineering, Nagaoka University of Technology, 1603-1, Kami-Tomioka, Nagaoka, Niigata 940-2188, Japan
[c] State Key Laboratory of Geo-hazard Prevention and Geo-environment Protection, Chengdu University of Technology, Chengdu, China
[d] Research Laboratory of Sedimentary Environment, Mineral and Water resources of Eastern Algeria, Larbi Tébessi University-Tebessa, Algeria
[e] Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran
[f] Department of Surface Mining, Hanoi University of Mining and Geology, 18 Vien st., Duc Thang ward, Bac Tu Liem dist., Hanoi, Vietnam
[g] Department of Civil and Environmental Engineering, University of Brasilia, Brazil
[h] Environmental Engineering and Earth Science Department, 445 Brackett Hall, Clemson University, Clemson, SC 29634, United States of America
[i] Faculty of Environment Earth Science, Hokkaido University, Sapporo 060-0810, Japan
[j] School of Energy and Mining Engineering, China University of Mining and Technology, Beijing 100083, China
[k] Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam
[l] Hokkaido Research Center of Geology (HRCG), Sapporo, Japan

HIGHLIGHTS

- Constructed a comprehensive inventory for 2018 Hokkaido coseismic landslides.
- Four different sampling techniques were tested for the predictive performance of LSM.
- DNN shows that different sampling strategies are less consequential in LSM.
- Applicability of results is validated with the 2015 Gorkha landslide cases.

GRAPHICAL ABSTRACT



DNN Model	AUC	ACC	Kappa	TP	TN	FP
Samples of Landslide Scarp	0.919	0.847	0.693	2099.3	2324	288.33
Centroid of Landslide Scarp	0.912	0.841	0.682	2057.6	2335.6	271.66
Samples of Landslide Body	0.913	0.842	0.683	2053	2344.6	277.33
Centroid of Landslide Body	0.904	0.829	0.657	2041	2288.3	309

ARTICLE INFO

Article history:
Received 17 December 2019
Received in revised form 13 February 2020
Accepted 13 February 2020

ABSTRACT

Predictive capability of landslide susceptibilities is assumed to be varied with different sampling techniques, such as (a) the landslide scarp centroid, (b) centroid of landslide body, (c) samples of the scrap region representing the scarp polygon, and (d) samples of the landslide body representing the entire landslide body. However, new advancements in statistical and machine learning algorithms continuously being updated the landslide

* Corresponding authors.
E-mail addresses: douj888@gmail.com (J. Dou), phamthaibinh2@duytan.edu.vn (B.T. Pham).
[1] These authors contributed equally.

https://doi.org/10.1016/j.scitotenv.2020.137320

Available online 15 February 2020

Editor: Damia Barcelo

Keywords:
Susceptibility
Landslide sampling strategies
Deep learning
Lidar DEM
M_w6.6 Hokkaido earthquake

susceptibility paradigm. This paper explores the predictive performance power of different sampling techniques in landslide susceptibility mapping in the wake of increased usage of artificial intelligence. We used logistic regression (LR), neural network (NNET), and deep learning neural network (DNN) model for testing and validation of the models. The tests were applied to the 2018 Hokkaido Earthquake affected areas using a set of 11 predictor variables (seismic, topographic, and hydrological). We found that the prediction rates are inconsequential with the DNN model irrespective of the sampling technique (AUC: 0.904 – 0.919). Whereas, testing with LR (AUC: 0.825 – 0.785) and NNET (AUC: 0.882 – 0.858) produces larger differences in the accuracies between the four datasets. Nonetheless, the highest success rates were obtained for samples within the landslide scarp area. The analogy was then validated with a published landslide inventory from the 2015 Gorkha earthquake. We, therefore, suggest that DNN models as an appropriate technique to increase the predictive performance of landslide susceptibilities if the landslide scarp and body are not characterized properly in an inventory.

1. Introduction

Landslides are typical of mountainous terrains, particularly in environments generally associated with earthquakes and rainstorms (Keefer, 1984; Li et al., 2020; Malamud et al., 2004; Wang et al., 2019; Yunus et al., 2020). Catastrophic landslides cause potentially adverse impacts in the downslope regions to people and their properties (Froude and Petley, 2018; Li et al., 2020a; Tien Bui et al., 2019). Therefore, a significant amount of research in the past two decades has been devoted to quantifying the controls of landslides and mapping their spatial distribution (Dou et al., 2015a; Gorum et al., 2011; Korup and Stolle, 2014). With latest advancements in statistical models and machine learning techniques that continuously being developed and updated, the spatial prediction of landslides in the form of vulnerability or landslide susceptibility maps (LSM) aiding in planning and mitigation strategies (Chang et al., 2019; Dou et al., 2019a; Santangelo et al., 2015).

Success rates of susceptibility maps are critical to effective landslide risk reduction. The predictive performance rates of landslide susceptibility methods rely mainly on the quality of input data (Hussin et al., 2016; Reichenbach et al., 2018). This includes the quality of landslide

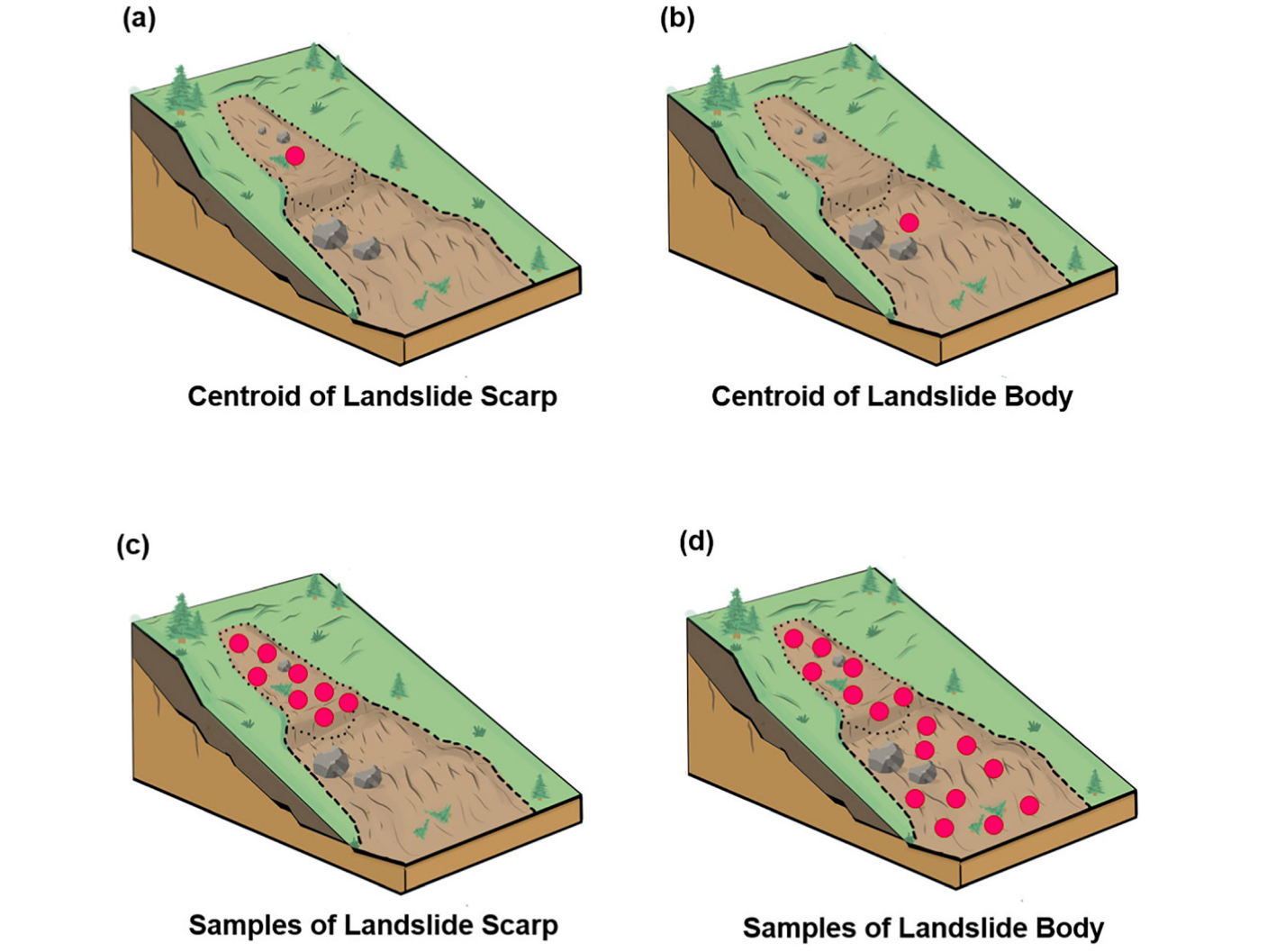


Fig. 1. Graphical sketch four types of landslide representative maps: (a) centroid of landslide scarp; (b) centroid of landslide body or boundary; (c) samples of landslide scarp; (d) samples of landslide body.

inventory, their completeness, and the input conditioning factor variables that feed into the model, most of which comes from a digital elevation model (DEM) (Dou et al., 2019b). While landslide conditioning factors are selected on the base of geoenvironmental characteristics of the study area in conjunction with the nature of landslide occurrence, landslide inventories are often obtained by field investigation or remote sensing approaches (Guzzetti et al., 2012; Reichenbach et al., 2018). The quality of DEM nowadays is found more or less accurate to ground truth with the introduction of LiDAR and unmanned aerial vehicle (UAV) based Structure from Motion (SfM) photogrammetric techniques (Chang et al., 2019; Sidle and Bogaard, 2016). Further, satellite-based DEMs from the TanDEM-X mission and ALOS mission are found of superior quality for performing geomorphological studies (Avtar et al., 2015; Tadono et al., 2016).

Contrarily, most of the landslide inventories are incomplete in the sense that they do not distinguish between landslide initiation zones and landslide runout zones (Korup and Stolle, 2014). A large number of studies in the recent past addressed that a reliable landslide inventory (i.e., quality of the landslide mapping method, positional accuracy, and the sampling strategy) is a vital to accurate susceptibility maps (Chang et al., 2019; Petschko et al., 2014; Steger et al., 2016). Because, in a geographic information system (GIS), the conditioning factor information is extracted accordingly with the positional precision and mapping/sampling criteria. Positional accuracy, though can be achieved with modern technologies such as differential Global Positioning Systems (GPS) assisted field surveys.

Notwithstanding, it is well documented in the literature that the selection of landslide pixels affects dramatically the effectiveness of landslide susceptibility modeling (Chang et al., 2019; Dou et al., 2015c; Pham et al., 2019). Landslide inventories for susceptibility map generation typically feed into the GIS in any one or combination of the following types (cf. Fig. 1): (a) the landslide scarp centroid, (b) centroid of landslide body, (c) samples of the scrap region representing the scarp polygon, and (d) samples of the landslide body representing the entire landslide body. However, until recently, only a few studies distinguished the initiation zone (scarp area) and the runout zone (part of landslide body), primarily because it is difficult to separate them from medium to low-resolution satellite imageries. Thus, it is required a thorough study on how the mapping quality, or in other words, the selection of landslide pixel/point samples can affect the landslide susceptibility modeling. Although a small number of studies in the recent past addressed the issue of sampling strategies and their effects in LSM, here we critically evaluating the differences in prediction accuracies of LSM obtained from the aforementioned four sampling strategies. We take up this issue, particularly because, we hypothesize that with the advancement in artificial intelligence (AI), the differences in prediction performance is more or less inconsequential between the sampling techniques, hence older inventories that haven't characterized for initiation and runout zones can be re-evaluated with an appropriate AI model for highly accurate LS maps.

To test our hypothesis, we selected the landslide affected areas during the September 2018 Mw 6.6 Hokkaido earthquake region, Japan.

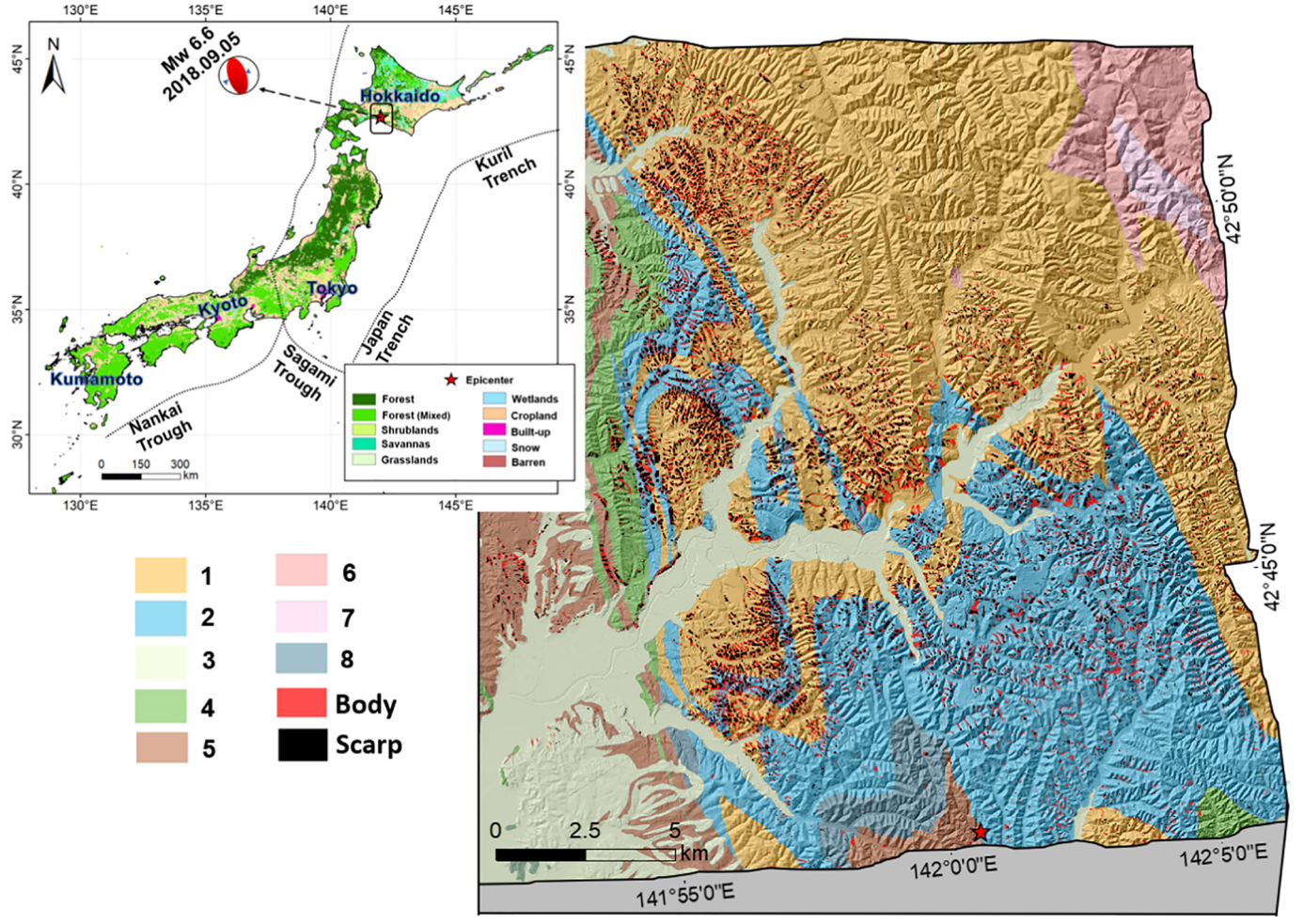


Fig. 2. Location of the case study showing earthquake-induced landslides (scarp and body), and major lithological units (1. Altered sandstone, mudstone, conglomerate, 2. shale, 3. sand and gravel, 4. conglomerate and sandstone, 5. sandstone and siltstone, 6. mudstone and sandstone, 7. sandy siltstone, and 8. siltstone and sandstone) Source: Geological Survey of Japan (GSJ): https://gbank.gsj.jp/datastore/.

Our rationale for selecting this case is because we obtained high-resolution pre- and post-earthquake DEM's and airborne imageries for the affected areas (please see Section 3), which aided in accurately producing landslide inventories with the distinction of initiation and runout zones. We selected three AI models to validate our hypothesis: logistic regression (LR) as a benchmark technique, artificial neural network (NNET), and deep learning neural network (DNN) as two very advanced machine learning models. DNN is a forked branch of NNET, therefore allows direct comparison, unlike the results from support vector machines (SVM) or tree-based models (e.g. Random Forest) can produce (Catani et al., 2013; Yunus et al., 2019). Moreover, there are only a handful number of studies carried using DL in the landslide field so far (Ghorbanzadeh et al., 2019; Huang and Xiang, 2018).

2. Overview of the study area

The study area is located in the Mw 6.6 earthquake struck region of Iburi-Tobu, southern Hokkaido, Japan, covering a part of the Yubari mountain range (Fig. 2). The affected areas are mainly located in Atsuma, Mukawa, and Abira towns of Hokkaido prefecture with comparatively low population densities (17 people/km^2). The epicenter is positioned at 42.686 N, 141.929 E, neighboring the Ishikari Lowland, the western culmination zone of the Hidaka Collision region. This lowland region has the presence of surficial faults that strikes between N20W and N20E; however, the present Iburi-Tobu Mw 6.6 earthquake has a deep focus (35 km) and has no observed surface rupture. This qualified the 2018 September 6th event an unusual deep-seated continental earthquake one that has a significant amount of coseismic landslides. The topographic features in the study area are characterized by gently undulating hills with elevations ranges between 100 m to 800 m above mean sea level. The mean slope angle in the study area is about 20°, and >70% of Terrain within the study area showcases slopes between 15° - 35° degrees. The surficial geology is comprised of blanketed volcanic deposits of several layers formed during the eruptions of mount Kuttara, Tarumae, Eniwa, and the Shikotsu caldera volcanoes (9000–40,000 years ago) (Yamagishi and Yamazaki, 2018). The deposits of these layers consist of unconsolidated pumice, scoria, and coarse-grained volcanic ash particles. Underlying these deposits, thickly bedded Quaternary and Neogene sedimentary rocks are present, predominantly consists of shale, and altered sandstone - conglomerate lithology.

The annual precipitation of the Hokkaido ranges between 1200 mm to 1800 mm. This rainfall is lower than what is observed in other regions of Japan (Yamagishi et al., 2002). The presence of large scale faulting and folding, weak, and low-consolidated sedimentary deposits, along with hilly terrain, makes slopes prone to sliding (Ayalew et al., 2011).

3. Modeling approach

The modeling approach used by this research can be summarized in Fig. 3. First, the earthquake-triggered landslides were mapped by interpreting aerial photographs, Lidar DEM, and ground truth to create a comprehensive landslide inventory for the study area. Subsequently, the landslide conditioning factor was examined. Thereafter, the LSM maps were produced by three data mining techniques, namely logistic regression, artificial neural network, and deep learning neural network. Eventually, the three models were investigated and validated for accuracy using the area under the receiver operating characteristic curve (AUC), and other evaluation matrices.

All analyses done by this research were conducted in Python (3.7.0) environment. Specifically, Tensorflow API (2.0.rc) was used exclusively to implement LR, NNET and DL models. The source code of this research

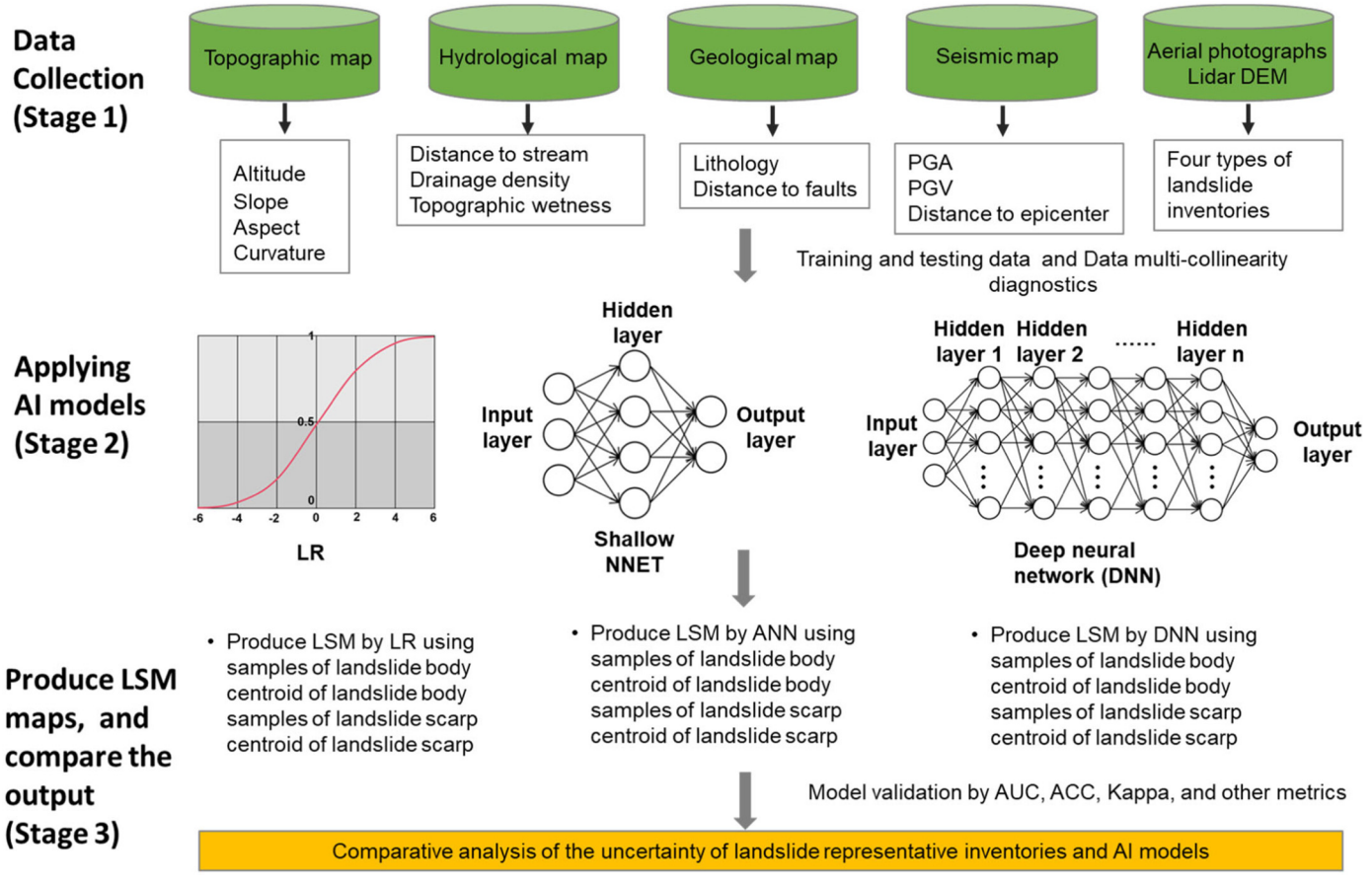


Fig. 3. The flowchart of this study

Table 1
Spatial data sources used in the study area.

Spatial database	Data	Type	Scale/resolution	Data source
Aerial photographs	Landslide	Raster	0.2 × 0.2 m	GSI
Lidar DEM	Landslide	Raster	2 × 2 m	Hokkaido Government
Topographic map	Morphometric factors	Raster	10 × 10 m	GSI
Geological map	Lithology	Polygon	1:200,000	GSJ
Tectonic factor	Faults	Polygon	1:200,000	GSJ
Seismic factor	PGA	Raster	10 × 10 m	GSJ
	PGV	Raster		
	Epicenter	Polygon		
Hydrological map	Stream Density Distance to stream network	Raster	10 × 10 m	GSI

is publically available online (https://www.github.com/aminevsaziz/). Software such as ArcGIS (10.4) and SAGA (4.0.1) were used for map compilation and production.

3.1. Landslide inventory

Preparing an accurate landslide inventory map is the first step in the precise prediction of landslide-prone areas (Guzzetti et al., 2012). For the 2018 Hokkaido earthquake affected areas, a complete landslide inventory maps were generated by manually interpreting 0.2 m aerial photographs, along with 2 m Lidar DEM for eliminating the shadow effect in the aerial photographs. The details of the data source and metadata used in this research are summarized in Table 1. The coseismal landslides were generally shallow types (Wang et al., 2019; Yamagishi and Yamazaki, 2018). Landslide initiation zone (scarp/crown area) cannot easily be distinguishable for shallow type landslides compared with the deep-seated ones; nevertheless, the high-resolution aerial images and elevation models aided us to separate the initiation zone (scarp/crown area) and runout zones for shallower landslides in our case area. We interpreted a total of 10,120 landslides from the images. It is worth mentioning that, only two coseismal inventories before this study (2015 Gorkha: Roback et al., 2017, and 2004 Niigata: Sekiguchi and Sato, 2006) have made the initiation zone and the accumulation area separated, making our inventory one of the unique, and qualified to prepare precise landslide susceptibility maps. Based on the inventory, four different sampling criteria were adopted to be used as input data in the susceptibility models. They are, (a) the point representing landslide centroid, (b) point representing landslide body centroid, (c) points representing the landslide scarp boundary, and (d) points representing the whole landslide body. All the points are derived from the scarp and landslide body polygon maps. Landside scarp here refers to the source of disposition/initiation, while landslide body is the boundary of the

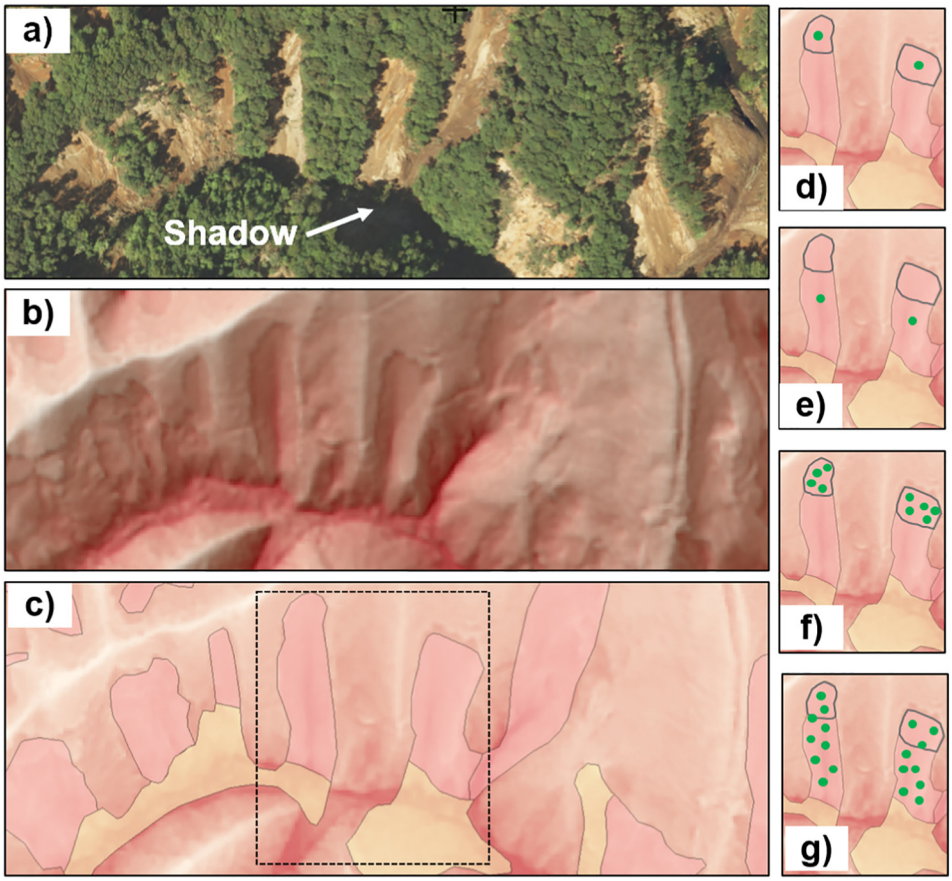


Fig. 4. Landslide inventory map and four different sampling strategies adopted in this study. a) aerial imagery showing cos-seismic landslides, b) i-value map generated from 2 m digital elevation model, c) inventory map showing landslide body (red color) and deposition zone (yellow color), d) point sampling from centroid of scarp area, e) point sampling from centroid of landslide body, f) multiple points from scarp area, and g) multiple points from landslide body.

landslide from source to runout area. Representative examples of the four sampling techniques shown in Fig. 4. The centroid points performance used in this research were converted from landslide polygons using ArcGIS 10.4 software, determined by the location of the center of the polygon of gravity.

The probability density distribution of coseismal landslides and area is plotted in Fig. 5.

3.2. Landslide conditioning factors

Selecting the appropriate landslide conditioning factors (LCF) is an essential task for the assessment of landslide susceptibility. To date, there are no established standard guidelines or universal agreement on the option of landslide conditioning factors. In this study, we have investigated earthquake-induced landslide that occurred during an Mw 6.6 seismic event in Hokkaido, Japan on September 6, 2018; thus, we selected seismic related parameters such as peak ground acceleration (PGA), peak ground velocity (PGV), and distance to the epicenter as the LCF. Besides, several other LCF, including topographical-hydrological-lithological variables such as altitude, slope, distance to stream network, stream density, Topographic Wetness Index (TWI), aspect, curvature, and lithology were selected for landslide susceptibility assessment in this study. To test the inter-associations among the independent variables, we executed a multi-collinearity analysis, because collinearity among the variables amplifies the discrepancy of regression analysis. The variance inflation factor (VIF) is used as an indicator of multicollinearity analysis (Khosravi et al., 2019). A VIF value >10 indicates significant multi-collinearity (O'Brien, 2007).

3.3. Logistic regression

The well-known multivariate analytical method, logistic regression (LR), has been more attended in the last decade for landslide susceptibility mapping (Budimir et al., 2015). Thus, LR is fit to be a benchmark model when a new model was developed and suggested (Chang et al., 2019). LR is based on a non-linear equation that is applicable for the binary variables such as 'presence' or 'absence' cases in the data. It can compute the weights for each conditioning factors as independent variables based on the binary dependent variable at a statistical confidence level (Shirzadi et al., 2012). The merits of LR include: (i) not require dataset to be normally distributed, (ii) dependent and independent variables can both be continuous or discrete, and (iii) it does not assume to having equal statistics variances among variables (Dou et al., 2019d). The weights are calculated by the LR method based on the following equations:

$$Z = c_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n \tag{1}$$

$$P_{LR} = 1 + e^{Z=c_0} + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n \Big/ e^{Z=c_0} + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n \tag{2}$$

where z is weighted linear combination of the independent variables that vary from $-\infty$ to $+\infty$, c_0 is the constant coefficient of the equation, α_1, α_2, ..., α_n are the coefficients of variables, x_1, x_2, ..., x_n are landslide conditioning factors, P_{LR} is the probability of landslides occurrence that either has been occurred (>0.5) or have been non-occurred (<0.5). The higher the P_{LR} for each pixel of the study area (closer to 1), the more probable a landslide pixel is to occur. The closer the P_{LR} to 0 will be, the less the probability of a pixel for landslide occurrence.

3.4. Artificial neural network

Artificial neural networks (NNETs) are a set of algorithms, modeled loosely after the human brain that is designed to recognize patterns. In fact, NNET was developed as a very crude approximation of nervous systems found in biological organisms (Swingler, 1996). The idea of NNETs is to transport information along a predefined path between "neurons". Neurons have the ability to add up information from multiple sources and they generally apply a non-linear transformation to this information in order to allow for more complex interactions. Contrary to some popular beliefs, the idea of NNETs is already very old (Hecht, 1987; Rumelhart et al., 1988). One of the first NNET to be invented was the perceptron. The perceptron was a very simple neural network with only one neuron and the Heaviside function as a non-linearity. In other words, the perceptron implements the following decision function:

$$f(x) = \begin{cases} 1, & \text{if } \omega x + \beta > 0 \\ 0, & x \le 0 \end{cases} \tag{3}$$

Its extension, the multilayer perceptron (MLP), can be regarded as the first "real" NNET, with multiple neurons connected in 2 layers, as shown in Fig. 6.

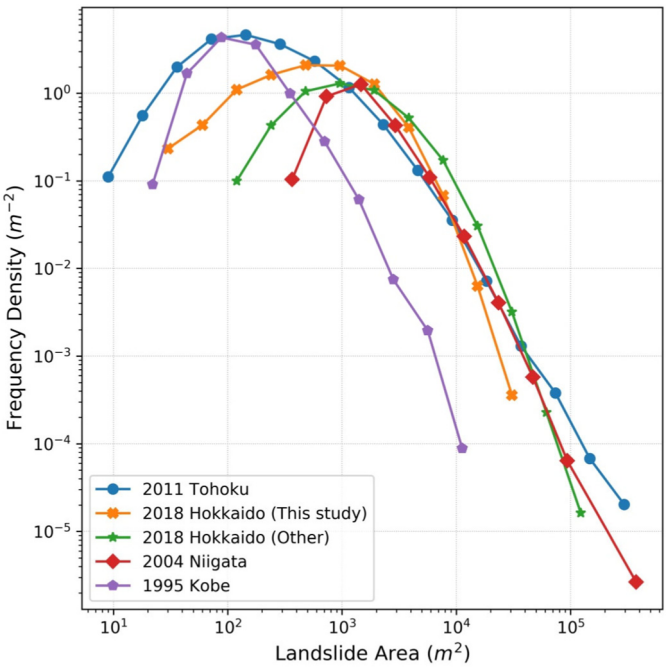


Fig. 5. Probability density distribution of coseismal landslides of the Hokkaido and other major earthquakes in Japan.

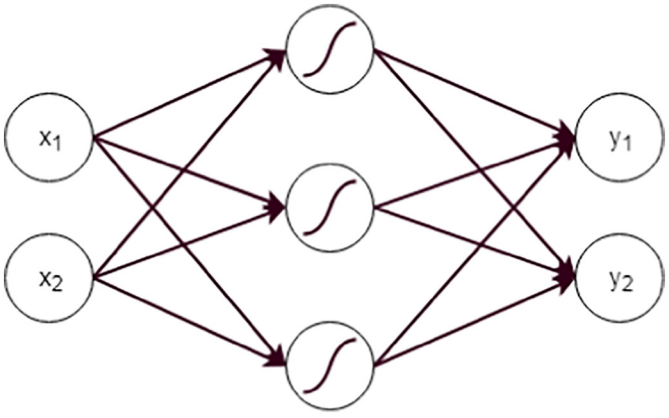


Fig. 6. Multilayer perceptron with sigma non-linearity: NNET consists of several layers; an input layer (X_n), a hidden layer (S) and an output layer (Y_n). Every layer consists of nodes, loosely modeled from neurons in the brain.

In general, MLP has much more expressive power than a single neuron. In fact, it can be shown that the multilayer perceptron is a universal function approximators (Paola and Schowengerdt, 1995). NNETs typically modify the multidimensional representation of input data by gradually applying (multi-)linear and non-linear transformations to its components and thus changing the representation up to a certain final result. The final result could be some classification or some other output signals. Most importantly, there is no notion of time in these networks. As far as the network is concerned, the output is generated instantaneously. Furthermore, the outputs of NNET neurons are continuously valued.

In NNET, we usually use a gradient descent technique to train the network. Gradient descent requires a large amount of data (the data has to be annotated for supervised learning) in order to determine the gradient of a predefined cost function and to converge to an (ideally global) minimum (Le et al., 2019b). Once the minimum is found, the network is considered trained and can be used for inference. Usually, no further training occurs at this stage. Multiple versions of gradient descent have been invented over time, with some of the more prominent being the Stochastic Gradient Descent (SGD), or Adaptive Moment Estimation (Adam). Thus, NNET method has been applied in the field of hazard mitigation because of good performance (Le et al., 2019a; LV et al., 2020; Shariati et al., 2019).

3.5. Deep learning

Deep learning (DL) is a self-teaching system, learning as it goes by filtering information through multiple hidden layers, in a similar way to humans. Technically speaking, deep learning neural networks (DNN) refers to "stacked artificial neural networks" (Bengio, 2009). This effectively distinguishes DNN from NNET (MLP) by their depth; that is, the number of hidden layers composed of. In fact, Deep learning is a phrase used for complex neural networks (Candel et al., 2016). The complexity is attributed to elaborate patterns of how information can flow throughout the model. Fig. 7 below illustrates an example of a DL network. The architecture has become more complex but the concept of DNN is still the same despite, there is now an increased number of hidden layers and nodes that integrate to estimate the output(s).

Essentially, DNN has a unique feature known as feature hierarchy, and it is a hierarchy of increasing complexity and abstraction. This makes DL networks capable of handling vast, high-dimensional data sets with billions of parameters that pass through nonlinear functions. Therefore, DNN can perform automatic feature extraction without human intervention, unlike most traditional machine-learning algorithms. Given that feature extraction is a task that can take teams of data scientist's years to accomplish, DNN is a way to circumvent the chokepoint of limited experts. It augments the powers of small data science teams, which by their nature do not scale (Deng and Yu, 2014).

NNETs have been invented over time, including the Convolutional Neural Network (CNN), thought to roughly emulating the human visual system, as well as the Recurrent Neural Network (RNN), used to interpret and generate sequential data such as text and video. Although the architectures of these networks are much more complicated than that of the MLP, they can be regarded as variations of the same idea. This can be seen for instance in the fact that all types of NNETs/DNN may use the same learning algorithms, e.g., gradient descent using backpropagation.

3.6. Conditioning factors analysis

In general, the conditioning factors used in landslide assessment frameworks are numerous, including terrain factors (obtained from a DEM) or factors derived by performing statistical certain operations on the terrain factor. However, the overabundance of data can undoubtedly lead to overfitting predictive model. As a result, three statistical tests have been implemented. The 1st and second statistical tests include a bivariate descriptor, Pearson Correlation matrix, and Variance Inflation Factors (VIF), respectively. The 3rd includes an entropy-based descriptor, Information Gain (IG).

The Pearson correlation and VIF statistical tests are common tests to detect high correlation among certain conditioning factors (i.e. multicollinearity) that lead to faulty modeling with erroneous system analysis (Merghadi et al., 2018). Specifically, these two tests assess the non-independence among conditioning factors. Generally, Pearson correlations focus on the covariance between each pair of factors divided by the product of their standard deviations (Eq. 4).

$$r_{x.y} = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sqrt{\sum_{k=1}^{n}(x_i - \overline{x})^2}} \times \frac{y_i - \overline{y}}{\sqrt{\sum_{k=1}^{n}(y_i - \overline{y})^2}} \quad (4)$$

where: n is the number of samples; x_i, y_i are conditioning factors indexed with i; $\overline{x}$ is the mean of x_i where: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, analogously same applies to $\overline{y}$.

On the contrary, VIF focus on the standard error variations of landslide conditioning factors, which imply the lower the standard errors, the lower multicollinearity risk, the safer the conditioning factor to implement (Eq. 5).

$$VIF = \frac{1}{1 - R^2} \quad (5)$$

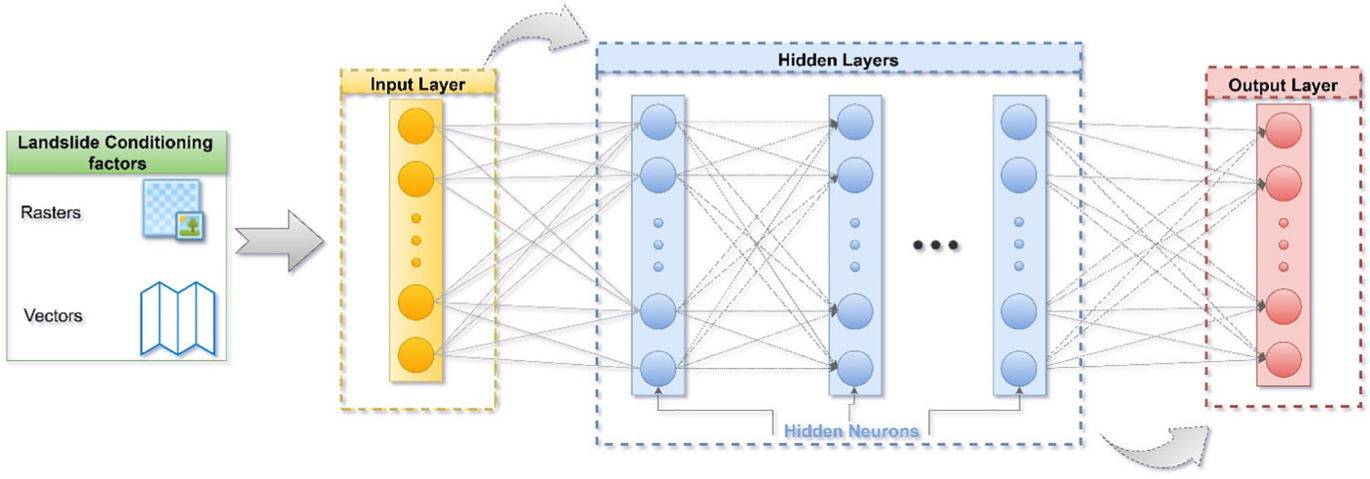


Fig. 7. An example of a deep neural network.

Table 2
Confusion matrix of landslide and non-landside occurrence conditions.

		Actual class		
		Landslide	Non-landslide	Total
Predicted class	Landslide	TP	FP	TP + FP
	Non-landslide	FN	TN	FN + TN
	Total	TP + FN	FP + FN	TP + FP + FN + TN

If the VIF value is >5 or correlation or correlation above 0.7, it indicates that there is a multicollinearity problem in the predisposing factors (Dormann et al., 2012; O'Brien, 2007).

The 3rd employed technique is Information Gain, defined as a reduction in entropy E(C) of a referent landslide inventory C (with j classes), due to the informational interference of a conditioning factor F (with n

Table 3
The multi-collinearity test for landslide conditioning factors using IG and VIF indexes.

Variable	The centroid of Landslide Body		The centroid of Landslide scarp		Samples of Landslide Body		Samples of Landslide Scarp	
	IG	VIF	IG	VIF	IG	VIF	IG	VIF
PGA	0.168	1.422	0.175	1.699	0.197	1.381	0.203	1.453
PGV	0.144	2.968	0.133	3.063	0.159	2.857	0.149	2.969
Epicenter	0.670	1.382	0.666	1.387	0.586	1.324	0.612	1.341
Altitude	0.678	3.208	0.674	2.861	0.615	3.285	0.633	3.015
Slope	0.679	1.567	0.676	1.591	0.630	1.499	0.646	1.535
Stream Distance	0.002	1.187	0.012	1.219	0.002	1.198	0.012	1.206
Stream Density	0.662	1.319	0.656	1.216	0.534	1.334	0.580	1.236
TWI	0.593	1.011	0.581	1.009	0.558	1.008	0.559	1.008
Aspect	0.679	1.023	0.673	1.025	0.626	1.017	0.643	1.010
Curvature	0.683	1.019	0.681	1.010	0.653	1.024	0.663	1.013
Lithology	0.132	1.189	0.135	1.177	0.122	1.146	0.142	1.166

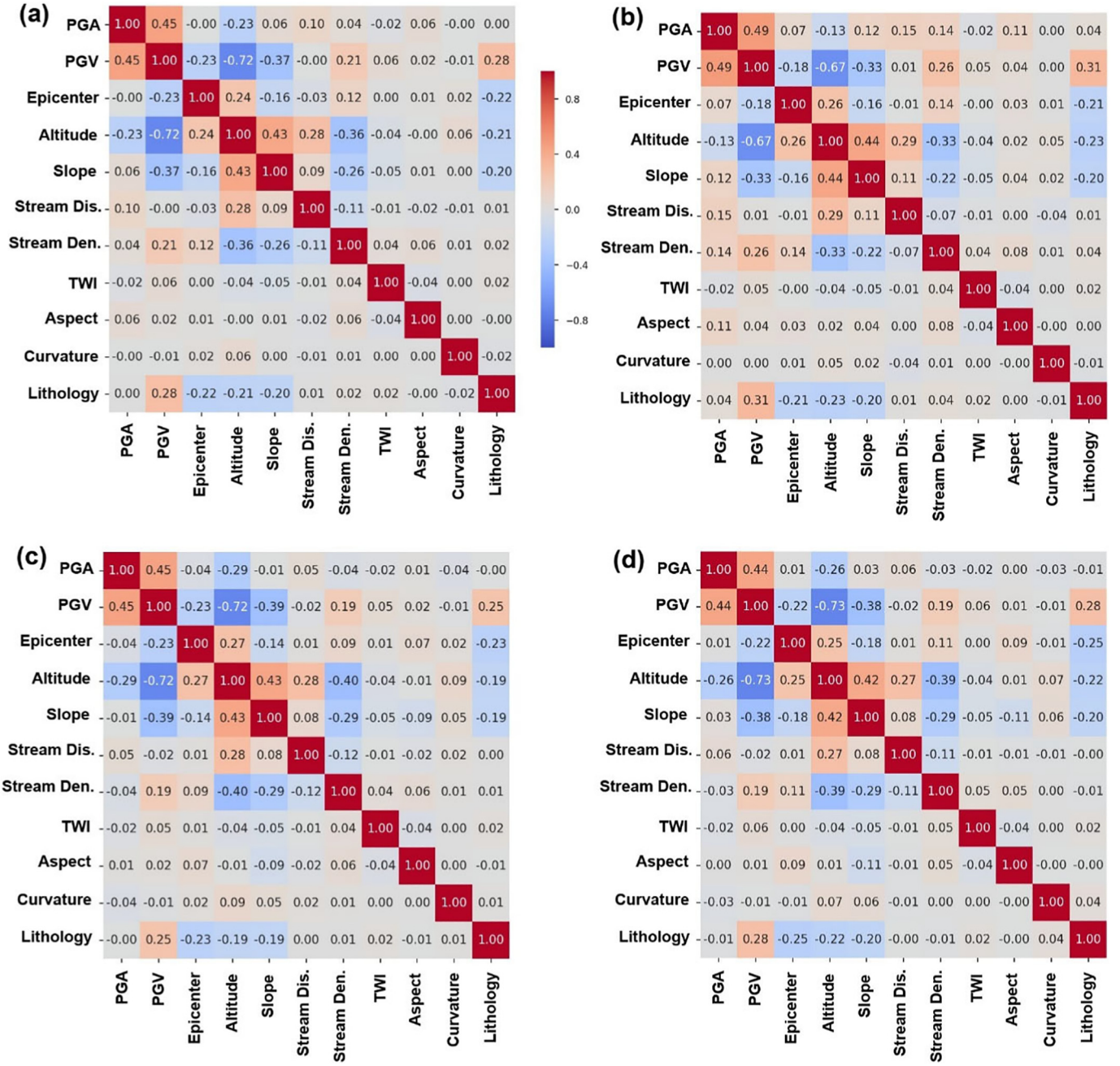
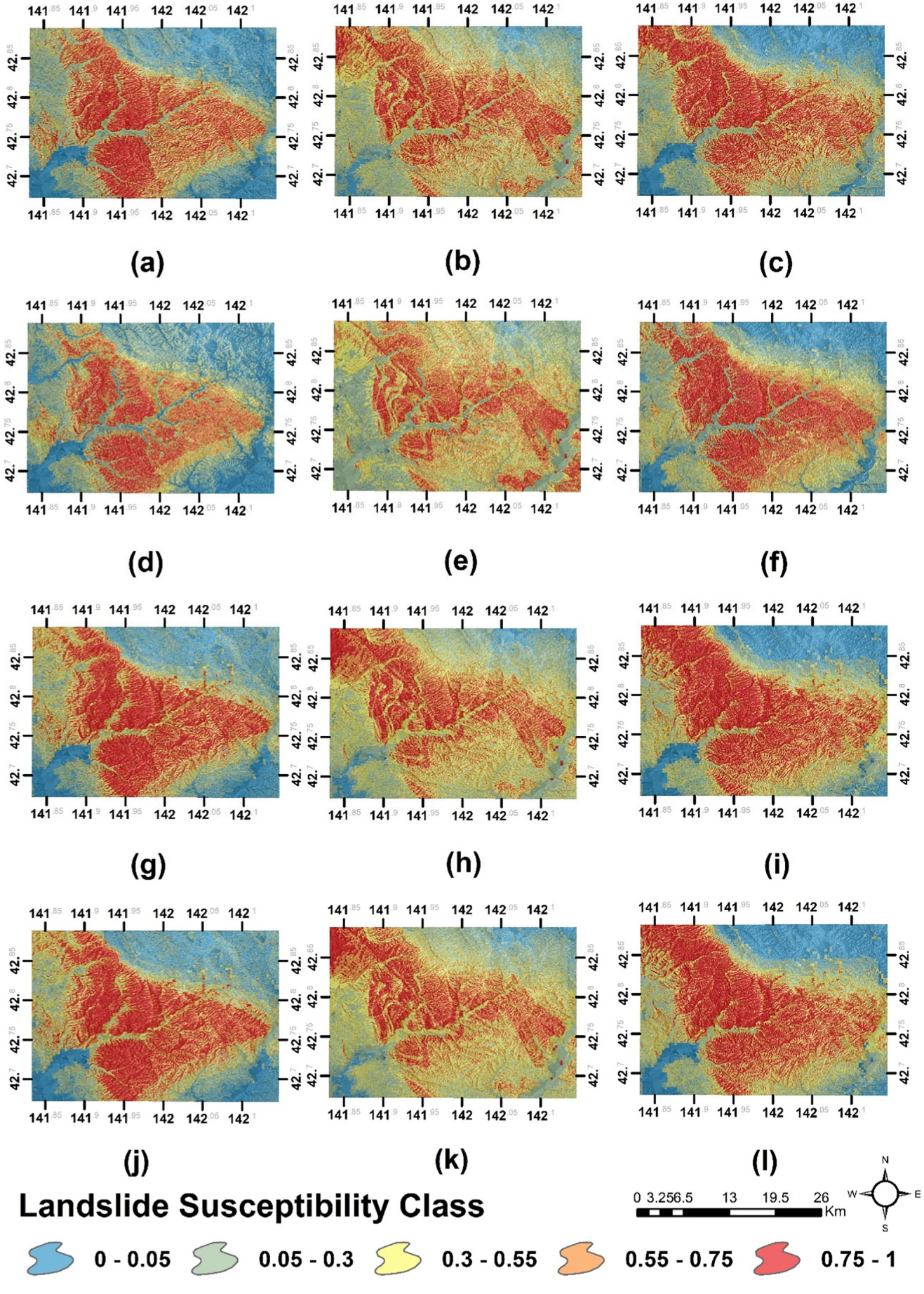


Fig. 8. The output results of the Pearson correlation matrix. **(a)** Samples of Landslide Scarp, **(b)** Centroid of Landslide Scarps, **(c)** Samples of Landslide Body, **(d)** Centroid of Landslide Body.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Landslide Susceptibility Class
0 - 0.05
0.05 - 0.3
0.3 - 0.55
0.55 - 0.75
0.75 - 1
0 3.256.5 13 19.5 26
Km
N
W
E
S

classes). Given the E(C) as a measure of homogeneity of C:

$$E(C) = -\sum_{i=1}^{j} \delta_i \log_2 \delta_i \quad (6)$$

where δ_i are a proportion of the i^{th} class values within the entire set, and introducing m factor classes with values $[v_1, v_2, \ldots v_m]$, the Information Gain IG(F) partitions the entropy by a factor of weighted expected entropy E(F,v).

$$IG(F) = E(C) - \sum_{v \in \{F|1,\ldots,m\}} \frac{|C_v|}{|C|} E(C_v) \quad (7)$$

The latter comes as summed entropy of C_v subsets of C, matched with the factor's class value v, and weighted by the subset proportion to C. Here, it should be mentioned that unlike Chi-Square statistic χ, this parameter allows preliminary ranking, since it disregards measure scales and units of conditioning factors (Gerath et al., 2006; Mitchell, 1997).

3.7. Model metrics

Developed LSM should be verified to check their ability for prediction accuracies (Pham et al., 2019; Tien Bui et al., 2019). Although there are some statistical metrics to evaluate the performance of a machine learning model, we in this study used accuracy (classification rate), kappa and area under the receiver operatic characteristic curve (AUC) metrics. The accuracy is computed using for possibility indices, including true positive (TP), true negative (TN), false positive (FP) and false-negative (FN). The TP and FP are the number of landslide pixels that correctly classified as landslide and non-landside pixels. However, TN and FN are a number of landslide pixels that incorrectly classified as landslide and non-landside pixels (Dou et al., 2019b; Pham et al., 2019). These possibilities can be expressed in the Table 2:

Accordingly, the accuracy, which is defined as the number of correctly classified landslide and non-landslide pixels, can be obtained using the four possibilities as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Another validation metric used in this study is the kappa index that shows the reliability of the landslide models (Bennett et al., 2013). A model is non-reliable if the value of the kappa is close to −1, while for values closer to 1 the model is reliable. It can be formulated as follows:

$$\text{Kappa} = \frac{P_a - P_{exp}}{1 - P_{exp}} \quad (9)$$

$$P_a = (TP + TN)/(TP + TN + FN + FP) \quad (10)$$

$$P_{exp} = \left((TP + FN)(TP + FP) + (FP + TN)(FN + TN)/\sqrt{(TP + TN + FN + FP)}\right) \quad (11)$$

According to Landis and Koch (1977), if the kappa index is <0, the model has a poor agreement between estimation (model) and observation (reality). However, the values of 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8 and 0.8–1, represents slight, fair, moderate, substantial, and almost conditions, respectively.

The other metric that has been used in this study is the receiver operating characteristics (ROC) curve. The ROC curve as a flexible and robust framework graphically evaluates the performance of a model. It is plotted based on the 100-specificity (false positive rate) and sensitivity (true positive rate) on the x-axis and y-axis, respectively. The specificity and sensitivity metrics can be computed according to the following equations:

$$Sensitivity = \frac{TN}{TN + FP}; y - axis \quad (12)$$

$$100 - Specificity = 100 - \left(\frac{TN}{TP + FN}\right); x - axis \quad (13)$$

The performance of a model by the ROC curve is quantitatively validated using the area under the curve (AUC) (Mason and Graham, 2002). It has been widely used as a popular and promising tool in the classification of performance. The model is perfect for the AUC to be close to 1; however, a value of 0.5 represents an inaccurate model (Youssef et al., 2015).

Table 4
Total landslide density of each susceptibility class in each susceptibility map.

Dataset	Model	Landslide density (%)				
		Very high	High	Moderate	Low	Very low
Samples of landslide scarp	DNN	88.581	7.612	1.211	2.076	0.519
	LR	78.201	18.512	2.422	0.865	0
	NNET	85.64	11.592	2.076	0.692	0
Centroid of landslide scarp	DNN	65.225	26.644	6.574	1.384	0.173
	LR	58.651	25.779	7.439	8.131	0
	NNET	65.571	14.533	8.997	8.997	1.903
Samples of landslide body	DNN	57.439	20.761	11.073	7.266	3.46
	LR	45.502	32.872	13.322	8.131	0.173
	NNET	53.46	26.298	12.976	5.536	1.73
Centroid of landslide body	DNN	55.200	21.800	8.834	9.505	4.599
	LR	43.263	33.911	11.083	10.37	1.312
	NNET	53.114	24.394	14.879	6.574	1.038

4. Results

4.1. Conditioning factors analysis

The output results of the Pearson correlation matrix (Fig. 8) show that certain variables are positively or negatively correlated with each other, yet they didn't exceed the allowed threshold of 0.7 for the correlation. For example, parameters such as PGA, PGV, Epicenter, Altitude, Slope, lithology and stream density, have good relationships with each other (up to 0.75). This indicates that using the aforementioned variables for earthquake-induced landslides are effective in LSM studies.

Table 3 shows the results of multicollinearity analysis and information gain contribution of all independent variables. The VIF results show that there is no significant collinearity among the variables, and the implemented models are safe to use since the VIF is less than the critical threshold of 5 (Dou et al., 2019a; Merghadi et al., 2018). The IG contribution of used variables is above 0.13 (except for distance to the stream with 0.002). Curvature shows the highest IG value, followed by variables such as aspect, slope, altitude, and distance to the epicenter. The IG for seismic variables (PGA and PGV) is found less prominent, probably because this study only considered point samples within the affected areas where they are more or less the same for both presence and absence cases.

4.2. Seismically-triggered landslides susceptibility assessment

The predicted landslide susceptibility maps using the implemented models produced raster grids in the format of susceptibility toward

Fig. 9. Landslide susceptibility maps produced for the study area (left column: DNN, central column: LR, right column: NNET): (***a-c***) Samples of Landslide Scarps, (***d-f***) Centroid of Landslide Body, (***g-i***) Samples of Landslide Body, (***j-l***) Centroid of Landslide Scarps.

Table 5
Total area extent covered by each susceptibility map.

Dataset	Model	Total area extent (%)				
		Very high	High	Moderate	Low	Very low
Samples of landslide scarp	DNN	19.004	19.014	9.057	11.987	40.938
	LR	4.451	20.835	19.824	29.557	25.332
	NNET	5.153	21.027	19.609	28.09	26.12
Centroid of landslide scarp	DNN	14.614	18.063	12.236	17.139	37.948
	LR	3.87	17.412	23.639	37.39	17.689
	NNET	2.509	29.042	17.11	23.184	28.154
Samples of landslide body	DNN	13.241	16.619	13.444	20.141	36.555
	LR	12.766	13.993	18.641	28.153	26.446
	NNET	6.032	24.619	17.719	23.351	28.279
Centroid of landslide body	DNN	3.25	27.941	17.56	28.494	22.756
	LR	4.451	20.835	19.824	29.557	25.332
	NNET	6.032	24.619	17.719	23.351	28.279

landsliding (Fig. 9). These probability raster maps are visual key to the overall quality of the implemented models.

The obtained LSM maps are shown in Fig. 9, where maps produced by DNN models show smooth maps, and the transition from each susceptibility class is unnoticeable. On the other hand, LR models produced overall bad LSM maps with a jittery gradient compared to maps generated by DNN and NNET. NNET-based maps were decent compared to both end of the spectrum (i.e. DNN and LR) therefore occupying the middle ground between DNN and LR based maps. However, in terms of the dataset used, maps generated based on scarps datasets (i.e. Fig. 9 (a–c) and (j–l)) produced overall better maps than body-based datasets. The results of the statistical distribution of susceptibility levels also show similar trends, given in Tables 4 and 5. Quantitatively, these results can be explained that a better model fit (i.e. generalization) of LSM is given by the scarp datasets than the other models.

4.3. Model comparison and validation

The performance results (Table 6 and Fig. 10) show that all models produced very good results (AUC > 0.785, Acc > 0.716 and kappa>0.433). In general, DNN outperforms the remaining models by considerable margin followed by NNET and, LR achieved the lowest performance out of the three implemented models. Moreover, landslide scraps achieved much better results than landslide bodies.

Nevertheless, the difference in performance results of the implemented models and datasets need to be assessed statistically if it is significant or not. Such a step is achieved by the non-parametric signed-rank test of Wilcoxon pairwise test using a significance threshold of <0.05.

The results of Wilcoxon test based on models pairwise (Table 7), demonstrate that in each datasets DNN effectively outperform the NNET, and NNET outperform LR model as the difference in performance between each pair of models reject the null hypothesis that suggests that there is no statistical significance in performance if p.value ≥0.05 and $-1.96 > z.value > +1.96$. On the other hand, the results of dataset pairwise (Table 8) indicate that: (1) landslide scraps datasets outperform landslide body datasets; and (2) sampling most of the landslide polygons are more effective than using a single polygon centroid.

5. Discussion

5.1. Susceptibility paradigms in deep learning era

Landslide inventory maps are quite important data for landslide assessment in the hilly terrains, especially the quality of landslide inventory maps (Chang et al., 2019; Guzzetti et al., 2012). Good quality of landslide inventory maps depends on the various factors, such as the scale, data acquired, quality of aerial photographs, etc. (Guzzetti et al., 2012). In this study, we collected various data sources to improve the quality of landslide inventory maps. There are difficulties in interpreting landslides in the shadow regions, and dense forested regions using ortho aerial photographs (Fig. 4a), however high resolution LiDAR elevation models when coupled with the red relief images (Fig. 4b) enhances the visualization and interpretation of topographic features (Görüm, 2019), providing accurate landslide mapping, creating the state of movement of landslides, and determining the initiation, runout zones. In terms of sampling, both the landslides with scarp and landslide bodies are effective for the prediction of LSM maps (Guzzetti et al., 2012). Some researchers only acquired point data to express the landslides for easiness or due to limited availability of high resolution imageries (Pham et al., 2019; Tien Bui et al., 2019), while some other researchers use landslide body polygon to symbolize the landslides,

Table 6
Performance results of the implemented models.

Dataset	Model		Metrics						
			AUC	ACC	Kappa	TP	TN	FP	FN
Samples of landslide scarps	DNN	Mean	0.919	0.847	0.693	2099.3	2324	288.33	513
		std	0.004	0.006	0.012	36.935	8.485	19.754	15.895
	LR	Mean	0.825	0.753	0.505	1853.6	2078.6	533.66	758.66
		std	0.003	0.002	0.004	31.899	26.03	13.474	12.284
	NNET	Mean	0.882	0.809	0.617	1988	2236.3	376	624.33
		std	0.005	0.005	0.011	44.967	19.754	8.485	21.638
Centroid of landslide scarps	DNN	Mean	0.912	0.841	0.682	2057.6	2335.6	271.66	559.66
		std	0.001	0.005	0.009	21.484	41.363	2.867	26.911
	LR	Mean	0.815	0.751	0.502	1809.3	2114	493.33	808
		std	0.003	0.005	0.01	28.194	13.589	38.003	14.306
	NNET	Mean	0.869	0.803	0.606	1935.6	2260.3	347	681.66
		std	0.002	0.002	0.002	36.971	44.063	20.928	22.425
Samples of landslide body	DNN	Mean	0.913	0.842	0.683	2053	2344.6	277.33	549.66
		std	0.007	0.008	0.017	44.594	2.867	28.825	21.297
	LR	Mean	0.811	0.737	0.475	1775.3	2077.3	544.66	827.33
		std	0.01	0.008	0.016	30.685	13.96	43.099	10.781
	NNET	Mean	0.859	0.787	0.573	1891.3	2219.3	402.66	711.33
		std	0.008	0.007	0.014	43.759	8.498	23.3	12.71
Centroid of landslide body	DNN	Mean	0.904	0.829	0.657	2041	2288.3	309	586.33
		std	0.002	0.003	0.006	65.335	54.786	7.071	10.781
	LR	Mean	0.785	0.716	0.433	1787.3	1955.6	641.66	840
		std	0.002	0.003	0.005	26.662	41.556	30.291	33.536
	NNET	Mean	0.864	0.797	0.593	1940	2222	375.33	687.33
		std	0.004	0.004	0.007	44.654	58.861	25.25	12.499

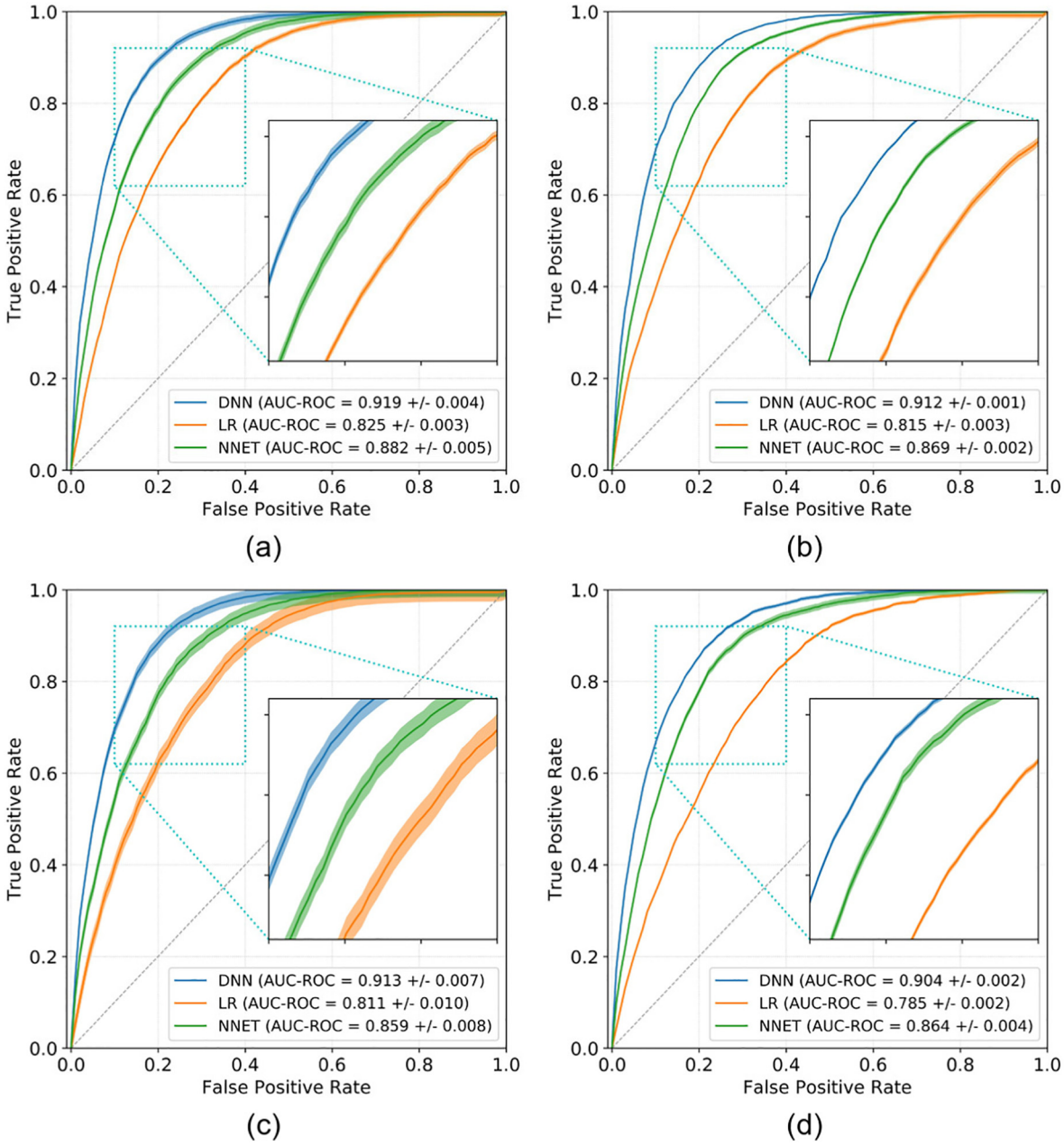


Fig. 10. ROC curves of the implemented models: **(a)** Samples of Landslide Scarp, **(b)** Centroid of Landslide Scarps, **(c)** Samples of Landslide Body, **(d)** Centroid of Landslide Body.

representing scarp/crown, deposition, etc. (Chang et al., 2019; Dou et al., 2019b). There are no agreements or universal rules on how to select the proper practices, or operational protocols, for the collection and updating landslide maps. The question on which landslide representative inventory maps for assessment of susceptibility is better, still keep the limitation of the complete discussion. In this study, we sampled four common types (cf. Section 3.1) to evaluate the effect uncertainty of landslide representative, namely, landslide scarp, the centroid of the scarp, landslide body, and the centroid of the body as shown in Figs. 1 and 4. In terms of accuracy of LSM, we noticed that sampling the scarp areas represent a better performance of the model than other methods, which are in agreement with a few previous types of research (Simon et al., 2013; Süzen and Doyuran, 2004). Though the centroid point sampling method is very quick, easy to the procedure and is automated for

Table 7
Models pairwise comparison on the implemented models using Wilcoxon pairwise test.

Pairwise	Dataset							
	The centroid of landslide body		The centroid of landslide scarp		Samples of landslide body		Samples of landslide scarp	
	p. Value	z. Value	p. Value	z. Value	p. Value	z. Value	p. Value	z. Value
DNN vs. LR	0.000	9.982	0.000	9.943	0.0002	7.165	0.000	4.666
DNN vs. NNET	0.002	3.574	0.006	8.921	0.0016	2.66	0.0004	2.458
LR vs. NNET	0.000	−5.995	0.000	−10.39	0.000	−4.471	0.0007	−5.839

Table 8
Datasets pairwise comparison on the implemented models using Wilcoxon pairwise test.

Pairwise	Model					
	DNN		LR		NNET	
	p. Value	z. Value	p. Value	z. Value	p. Value	z. Value
Samples of landslide scarp vs. centroid of landslide scarp	0.001	2.376	0.000	3.349	0.000	2.165
Samples of landslide scarp vs. samples of landslide body	0.000	4.554	0.001	6.829	0.000	5.868
Samples of landslide scarp vs. centroid of landslide body	0.000	5.235	0.000	7.368	0.000	6.276
Centroid of landslide scarp vs. samples of landslide body	0.002	2.289	0.001	3. 546	0.000	2.723
Centroid of landslide scarp vs. centroid of landslide body	0.000	3.645	0.000	5.753	0.000	4.786
Samples of landslide body vs. centroid of landslide body	0.000	4.998	0.000	6.545	0.000	4.471

saving the time, the method to allocate a point inside a polygon depends exclusively on the center of gravity of the polygon, henceforth resulting in lower accuracy. Scarp stands for unstable areas and undisturbed morphological conditions. As the landslide point pixel sampling strategy uses a single centroid vs the entire polygon, the success and prediction of the landslide significantly increased with the polygon sampling strategy. Similarly, the scarp polygon of the landslide gives a high success rate in predicting when compared with landslide bodies. Choosing the approximate landslide inventory maps, depending on the purpose, such as landslide body polygon is suitable for computing the landslide area, the point is straightforward to interpret landslide location and save time, landslide scarp is better of assessment of susceptibility maps.

An interesting observation in our analysis is that with DNN, the AUC values of the four different approaches yield more or less similar values within a range of 0.015 (Table 6). This suggests that irrespective of the sampling strategy, a deep learning algorithm accurately predicts the landslide susceptibilities with only a marginal difference. However, the difference in accuracies noticed in other traditional models (i.e., LR and ANN) with multiple sampling techniques yields a larger difference in AUC value. This implies that DNN can successfully be applied for the cases where inventories are available for the past, but no separation is made between the scarp and body, or even for the point inventories. Indeed, it is true that except for 2015 Gorkha (Roback et al., 2017) and 2004 Niigata (Sekiguchi and Sato, 2006) earthquake datasets, all other published co-seismic landslide inventories are only having their landslide bodies demarcated.

The advantages of DNN with a gradient descent approach include fast and simple implementation and relatively fast convergence as compared to algorithms such as Genetic Algorithms (Arifovic and Gencay, 2001; Dou et al., 2015b). While the limitation lies in the fact that in order to apply the gradient descent algorithms in DNN, the network has to be static, i.e., the connections between neurons may change in strength but not in number. Therefore, during the final inference stage, the computational load remains the same regardless of the complexity of the real problem. Another issue with DNN is its tendency to over-fit the training data. Overfitting causes the network to have an excellent performance on training data and bad performance on test data. For complex problems such as computer vision problems that require networks with millions of parameters, the only way to alleviate this issue is to provide more training data. This can limit the applicability of DL for problems with small amounts of available data. Nevertheless, there exist numerous landslide inventories in the published domain with either point data or the only body is demarcated, but the assessments with DNN could change the current susceptibility paradigms with such datasets.

5.2. Validation of DNN results with Gorkha 2015 inventory

We validated our claim on the results about the non-necessity of multiple sampling strategies in the future with DNN models as the AUC yields similar output, and their applicability to past and future landslide susceptibility assessments. For this purpose, we used the published 2015 Gorkha earthquake-induced landslide inventory (Roback

Table 9
Results for Nepal Case Study using multiple sampling techniques and susceptibility models.

Dataset	Model		Metric						
			AUC	ACC	Kappa	TP	TN	FP	FN
Samples of landslide scarps	DNN	Mean	0.918	0.841	0.681	1439.330	1468.000	240.000	311.667
		std	0.006	0.006	0.013	30.576	14.855	22.554	7.364
	LR	Mean	0.845	0.775	0.549	1334.330	1344.670	363.333	416.667
		std	0.005	0.006	0.012	45.346	31.668	14.079	11.025
	NNET	Mean	0.915	0.839	0.678	1435.330	1467.000	241.000	315.667
		std	0.005	0.003	0.006	35.462	25.807	8.602	5.907
Centroid of landslide scarps	DNN	Mean	0.899	0.818	0.635	1382.330	1445.670	262.333	368.667
		std	0.007	0.005	0.010	11.146	23.921	23.099	35.349
	LR	Mean	0.789	0.726	0.453	1223.670	1287.670	420.333	527.333
		std	0.017	0.023	0.046	41.652	43.053	71.853	7.717
	NNET	Mean	0.893	0.813	0.626	1383.000	1428.000	280.000	368.000
		std	0.008	0.009	0.018	6.377	27.580	18.019	39.908
Samples of landslide body	DNN	Mean	0.897	0.814	0.627	1333.670	1480.330	270.667	374.333
		std	0.001	0.002	0.004	29.781	29.959	17.172	24.088
	LR	Mean	0.824	0.760	0.520	1237.670	1392.000	359.000	470.333
		std	0.005	0.004	0.008	33.230	27.653	16.391	29.488
	NNET	Mean	0.891	0.808	0.615	1328.330	1465.000	286.000	379.667
		std	0.002	0.003	0.006	26.637	28.083	28.717	18.927
Centroid of landslide body	DNN	Mean	0.896	0.819	0.638	1339.000	1494.000	257.000	369.000
		std	0.004	0.003	0.006	33.476	32.752	10.985	1.633
	LR	Mean	0.802	0.741	0.482	1197.670	1365.670	385.333	510.333
		std	0.007	0.008	0.016	11.898	36.573	26.612	22.603
	NNET	Mean	0.890	0.815	0.629	1344.000	1474.000	277.000	364.000
		std	0.006	0.005	0.011	30.232	33.025	13.140	7.483

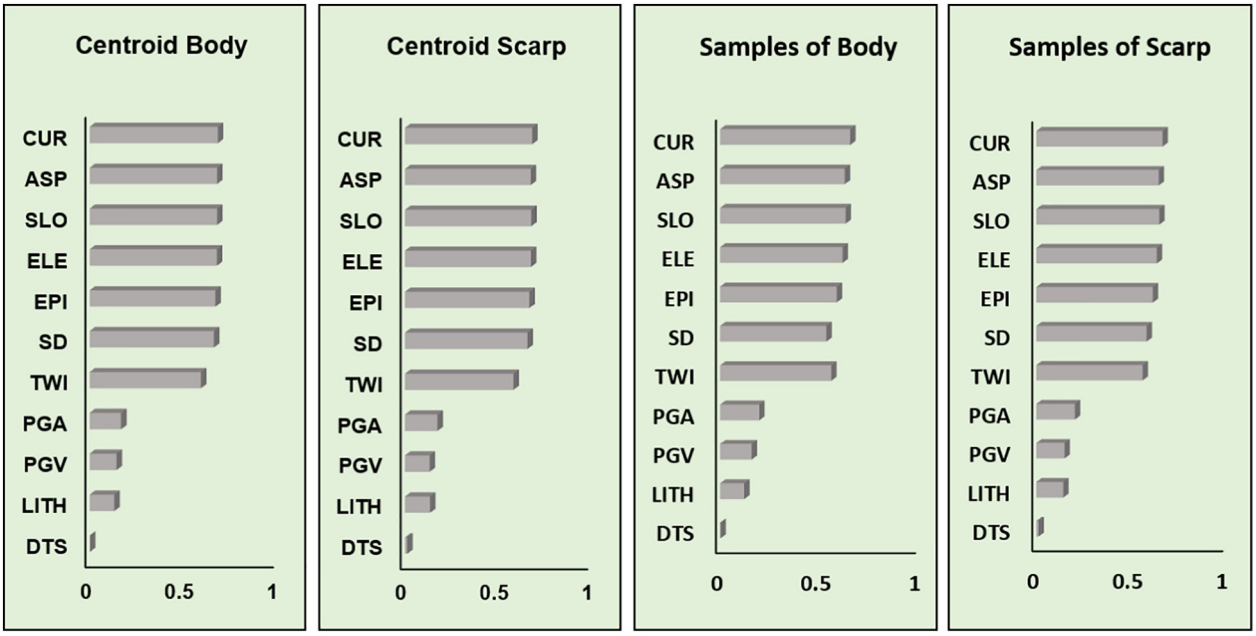


Fig. 11. Landslide conditioning factors rank importance using the IGR indicator.

et al., 2017). Similar to our inventory, Roback et al. (2017) also mapped both scarp and landslide bodies, thus helped for a direct comparison. We used similar LCF and an ALOS World 3D DEM for preparing the results. The results of the Gorkha cases are shown in Table 9.

The results obtained for Gorkha case also shows that DNN outperformed all other susceptibility models, and the difference in four different sampling techniques is less consequential; i.e., the AUC variation is only less than or equal to 0.022. This confirms the validity of our hypothesis.

5.3. What controls the coseismic landslides

The assessment of the effect of the geo-environmental landslide conditioning factors to the methods has absorbed a lot of attention in the landslide susceptibility and has been debated in the numerous studies formerly (Dou et al., 2019a, 2019c; Pham et al., 2019; Tien Bui et al., 2019). The result of the assessment of input factors importance related to landsliding using IGR is shown in Table 3 and Fig. 11. It is worth mentioning that the IGR ranking remains the same despite the change in sampling approaches, indicating the robustness of the modelled results. The results indicate that the topographical factors mainly influence the coseismal landslides. Both convex and concave curvature has a high significance in our study area. This can be attributed to the seismic amplification effect as shown by Maufroy et al. (2014), where they suggest that topographic curvature is correlated with the site amplification factor since they can capture the frequency-dependent variabilities. Wang et al. (2019), in their studies by analyzing coseismal landslides density and factor analysis reported the high correspondence of slope and

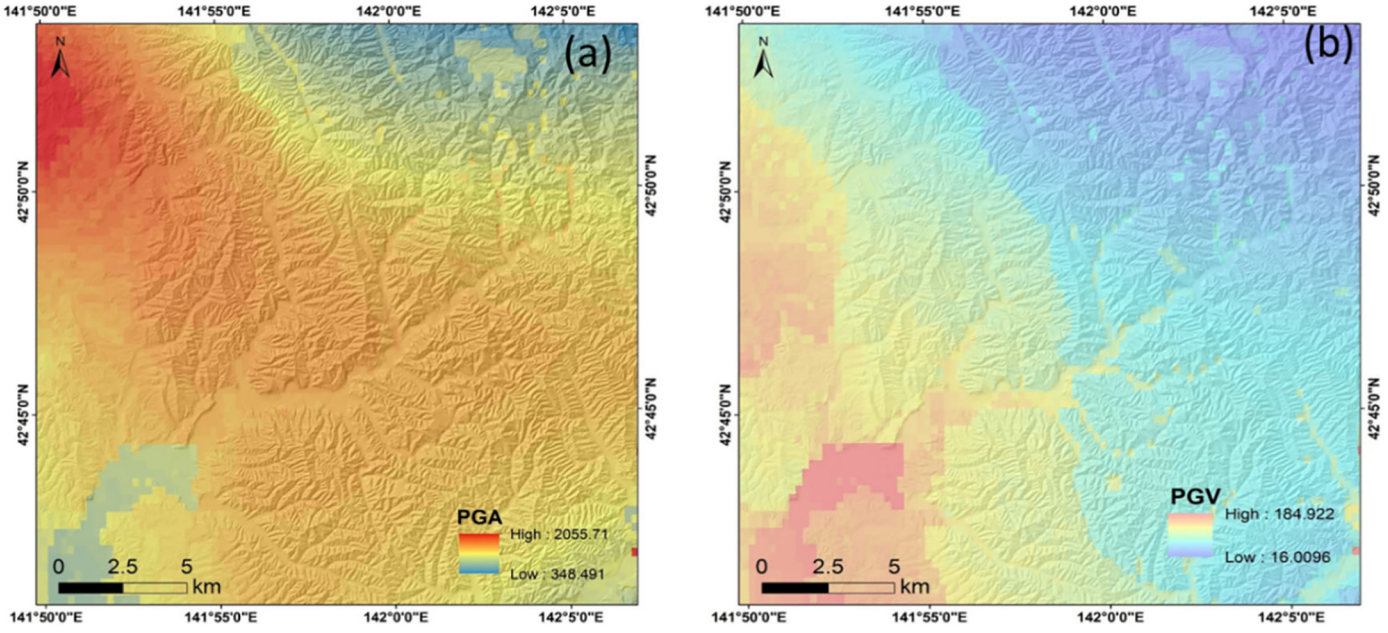


Fig. 12. Interpolated map of seismo-tectonic factors for Hokkaido earthquake: **(a)** PGA; **(b)** PGV.

altitude factors in the occurrences of landslides, which also agree with our IGR results. As for the aspect factor is considered, we noticed that the southeast (SE) aspect is parallel to the strike of the seismogenic source estimated from the moment tensor solutions proposed by the United States Geological Survey (USGS). Here in this study, the importance of the seismic factors is not well represented by IGR because these factors were made from interpolation of a few points, away from the affected areas. Moreover, both the landslide presence and absence samplings points are taken from a similar seismic zone as shown in Fig. 12. In addition, the bedrock lithology also does not contribute much to the model and observed results, because the landslides are mostly shallow and, the surficial geology is mostly covered by thick layers (>8 m) of weathered tephra (Osanai et al., 2019), deposited during the eruptions of mount Kuttara, Tarumae, Eniwa and the Shikotsu caldera volcanoes (9000–40,000 years ago) (Yamagishi and Yamazaki, 2018).

6. Conclusions

A detailed and comprehensive inventory of over 10,000 individual landslides perceived in the aftershock September 6th, 2018 M_w 6.6 Hokkaido earthquake has been compiled with the help of aerial photographs and Lidar DEM, four types of samples were extracted from the polygon shapes, i.e., samples of landslide scarp, centroid of scarp, samples of landslide body and centroid of body, respectively. Three machine learning models, namely, DNN, LR, and NNET were used to carry out the landslide susceptibility assessments. The AUC values of models are over 0.785 with good results. We found that the proposed DNN method has the best performance (AUC = 0.919), following NNET (AUC = 0.882), and LR (AUC = 0.825). In terms of LSM performance, the order of predictive power is in the following order: i.e., landslide scarp > landslide body > centroid of scarp > centroid of the body. We observed that with DNN, the accuracies between the four types of samples are more or less inconsequential. This implies a high potential for deep learning in future assessments. However, the aerial percentages and landslide density in five susceptibility classes between the four sampling techniques are found different. Therefore, the conclusion drawn from our study should be used with caution and required to be tested for more cases. Moreover, the model should be fine-tuned to address this issue.

The results of IGR show that topographic amplification in the steep slopes of the low-elevated terrain leads to enormous landslides. We hope the finding of this work may assist landslide investigators to determine the appropriateness of the type of landslide data and models because it can affect the accuracy of the final landslide susceptibility maps. This can be useful for developing appropriate hazard management practices. In the future, long-term special effects on the environment should be considered, such as the following snowmelt and heavy rainfall may exaggerate the unstable surface slope failure and cause serious secondary hazard disasters to occur, the dynamic landslide datasets and more conditioning factors should be considered and updated.

CRediT authorship contribution statement

Jie Dou: Conceptualization, Investigation, Formal analysis, Writing - review & editing. **Ali P. Yunus:** Investigation, Formal analysis, Writing - review & editing. **AbdelAziz Merghadi:** Investigation, Formal analysis and Writing. **Ataollah Shirzadi:** Writing - review & editing. **Hoang Nguyen:** Writing - review & editing. **Yawar Hussain:** Investigation, Writing - review & editing. **Ram Avtar:** Writing - review & editing. **Yulong Chen:** Writing - review & editing. **Binh Thai Pham:** Writing - review & editing. **Hiromitsu Yamagishi:** Writing - review & editing.

Declaration of competing interest

No conflict of interest exists.

We wish to confirm that there are no known conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

We express our sincere acknowledgements to Damià Barceló, Editor- in-chief of the Journal of Science of the Total Environment and the anonymous reviewer. This work has been supported by CAS Pioneer Hundred Talents Program, JSPS Program, and the opening fund from State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University (Grant No. SKHL1903).

References

Arifovic, J., Gencay, R., 2001. Using Genetic Algorithms to Select Architecture of a Feedforward Artificial Neural Network. 289 pp. 574–594.
Avtar, R., Yunus, A.P., Kraines, S., Yamamuro, M., 2015. Evaluation of DEM generation based on Interferometric SAR using TanDEM-X data in Tokyo. Phys. Chem. Earth https://doi.org/10.1016/j.pce.2015.07.007.
Ayalew, L., Kasahara, M., Yamagishi, H., 2011. The spatial correlation between earthquakes and landslides in Hokkaido (Japan), a GIS-based analysis of the past and the future. Landslides https://doi.org/10.1007/s10346-011-0262-z.
Bengio, Y., 2009. Learning deep architectures for AI. Found. Trends Mach. Learn. 2, 1–27. https://doi.org/10.1561/2200000006.
Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., 2013. Characterising performance of environmental models. Environ. Model. Softw. Softw. 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.
Budimir, M.E.A., Atkinson, P.M., Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides https://doi.org/10.1007/s10346-014-0550-5.
Candel, A., Parmar, V., LeDell, E., Arora, A., 2016. Deep Learning With H2O. H2O. ai Inc.
Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat. Hazards Earth Syst. Sci. 13, 2815–2831. https://doi.org/10.5194/nhess-13-2815-2013.
Chang, K.-T., Merghadi, A., Yunus, A.P., Pham, B.T., Dou, J., 2019. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. Sci. Rep. 9, 12296. https://doi.org/10.1038/s41598-019-48773-2.
Deng, L., Yu, D., 2014. Deep learning: methods and applications. Found. Trends® Signal Process. 7, pp. 197–387.
Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.) 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.
Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015a. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10, e0133262. https://doi.org/10.1371/journal.pone.0133262.
Dou, J., Chang, K.T., Chen, S., Yunus, A.P., Liu, J.K., Xia, H., Zhu, Z., 2015b. Automatic case-based reasoning approach for landslide detection: integration of object-oriented image analysis and a genetic algorithm. Remote Sens. 7, 4318–4342. https://doi.org/10.3390/rs70404318.
Dou, J., Yamagishi, H., Pourghasemi, H.R., Yunus, A.P., Song, X., Xu, Y., Zhu, Z., 2015c. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan. Nat. Hazards 78, 1749–1776. https://doi.org/10.1007/s11069-015-1799-2.
Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Han, Z., Pham, B.T., 2019a. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides https://doi.org/10.1007/s10346-019-01286-5.
Dou, J., Yunus, A.P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Khosravi, K., Yang, Y., Pham, B.T., 2019b. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci. Total Environ. 662, 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221.
Dou, J., Yunus, A.P., Tien Bui, D., Sahana, M., Chen, C.-W., Zhu, Z., Wang, W., Pham, B.T., 2019c. Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. Remote Sens. 11, 638. https://doi.org/10.3390/rs11060638.
Dou, J., Yunus, A.P., Xu, Y., Zhu, Z., Chen, C.-W., Sahana, M., Khosravi, K., Yang, Y., Pham, B.T., 2019d. Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. Nat. Hazards 97, 579–609. https://doi.org/10.1007/s11069-019-03659-4.
Froude, M.J., Petley, D.N., 2018. Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. https://doi.org/10.5194/nhess-18-2161-2018.
Gerath, R., Jakob, M., Mitchell, P., VanDine, D., Finn, L., Gillespie, D., Kuan, S., Naesgaard, E., Patrick, B., Skermer, N., 2006. Guidelines for Legislated Landslide Assessments for Proposed Residential Development in British Columbia. Assoc. Prof. Eng. Geosci. Br, Columbia, Vancouver, BC Google Sch.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J., 2019. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sens. https://doi.org/10.3390/rs11020196.
Görüm, T., 2019. Landslide recognition and mapping in a mixed forest environment from airborne {LiDAR} data. Eng. Geol. 258, 105155.
Gorum, T., Fan, X., van Westen, C.J., Huang, R.Q., Xu, Q., Tang, C., Wang, G., 2011. Distribution pattern of earthquake-induced landslides triggered by the 12 may 2008 Wenchuan earthquake. Geomorphology 133, 152–167. https://doi.org/10.1016/j.geomorph.2010.12.030.
Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K., 2012. Landslide inventory maps: new tools for an old problem. Earth-Science Rev 112, 42–66. https://doi.org/10.1016/j.earscirev.2012.02.001.
Hecht, N., 1987. IEEE first annual international conference on neural networks San Diego, California June 21–24, 1987. IEEE Expert 2, 14. https://doi.org/10.1109/MEX.1987.4307059.
Huang, L., Xiang, L. yang, 2018. Method for meteorological early warning of precipitation-induced landslides based on deep neural network. Neural. Process. Lett. https://doi.org/10.1007/s11063-017-9778-0.
Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2016. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology 253, 508–523. https://doi.org/10.1016/j.geomorph.2015.10.030.
Keefer, D.K., 1984. Landslides Caused by Earthquakes. pp. 406–421.
Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H.L., Hong, H., Chapi, K., Prakash, I., 2019. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J. Hydrol. 573, 311–323. https://doi.org/10.1016/j.jhydrol.2019.03.073.
Korup, O., Stolle, A., 2014. Landslide prediction from machine learning. Geol. Today https://doi.org/10.1111/gto.12034.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 159–174.
Le, L.T., Nguyen, H., Dou, J., Zhou, J., 2019a. A comparative study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings' energy efficiency for smart city planning. Appl. Sci. 9, 2630. https://doi.org/10.3390/app9132630.
Le, L.T., Nguyen, H., Jian, Z., Dou, J., 2019b. Estimating the heating load of energy efficiency of buildings for smart city planning using a novel artificial intelligence technique PSO-XGBoost. Appl. Sci. 9 (13), 2630. https://doi.org/10.3390/app9132630.
Li, H., Xu, Y., Zhou, J., Wang, X., Yamagishi, H., Dou, J., 2020. Preliminary analyses of a catastrophic landslide occurred on July 23, 2019, in Guizhou Province, China. Landslides https://doi.org/10.1007/s10346-019-01334-0.
Li, Y., Liu, X., Han, Z., Dou, J., 2020a. Spatial proximity-based geographically weighted regression model for landslide susceptibility assessment: a case study of Qingchuan area, China. Appl. Sci. 10, 1107. https://doi.org/10.3390/app10031107.
LV, Y., Le, Q., Bui, H.-B., Bui, X., Nguyen, H., Nguyen-Thoi, T., Dou, J., Song, X., 2020. A comparative study of different machine learning algorithms in predicting the content of ilmenite in titanium placer. Appl. Sci. 10, 635. https://doi.org/10.3390/app10020635.
Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslides, earthquakes, and erosion. Earth Planet. Sci. Lett. 229, 45–59. https://doi.org/10.1016/j.epsl.2004.10.018.
Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics ({ROC}) and relative operating levels ({ROL}) curves: {statistical} significance and interpretation. Q. J. R. Meteorol. Soc. A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr. 128, 2145–2166.
Maufroy, E., Cruz-Atienza, V.M., Cotton, F., Gaffet, S., 2014. Frequency-scaled curvature as a proxy for topographic site-effect amplification and ground-motion variability. Bull. Seismol. Soc. Am. 105, 354–367.
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-Information 7. https://doi.org/doi:https://doi.org/10.3390/ijgi7070268.
Mitchell, T.M., 1997. Machine Learning. WCB.
O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41, 673–690. https://doi.org/10.1007/s11135-006-9018-6.
Osanai, N., Yamada, T., Hayashi, S. ichiro, Kastura, S., Furuichi, T., Yanai, S., Murakami, Y., Miyazaki, T., Tanioka, Y., Takiguchi, S., Miyazaki, M., 2019. Characteristics of landslides caused by the 2018 Hokkaido eastern Iburi earthquake. Landslides, 1517–1528 https://doi.org/10.1007/s10346-019-01206-7.
Paola, J.D., Schowengerdt, R.A., 1995. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. Int. J. Remote Sens. https://doi.org/10.1080/01431169508954607.
Petschko, H., Brenning, A., Bell, R., Goetz, J., Glade, T., 2014. Assessing the quality of landslide susceptibility maps - case study Lower Austria. Nat. Hazards Earth Syst. Sci. 14, 95–118. https://doi.org/10.5194/nhess-14-95-2014.
Pham, B.T., Prakash, I., Dou, J., Singh, S.K., Trinh, P.T., Tran, H.T., Le, T.M., Van Phong, T., Khoi, D.K., Shirzadi, A., Bui, D.T., 2019. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. Geocarto Int, 1–25 https://doi.org/10.1080/10106049.2018.1559885.
Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Rev https://doi.org/10.1016/j.earscirev.2018.03.001.
Roback, K., Clark, M.K., West, A.J., Zekkos, D., Li, G., Gallen, S.F., Champlain, D., Godt, J.W., 2017. Map Data of Landslides Triggered by the 25 April 2015 Mw 7.8 Gorkha, Nepal Earthquake: U.S. Geological Survey Data Release. https://doi.org/doi:https://doi.org/10.5066/F7DZ06F9.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning internal representations by error propagation. In: Collins, A., Smith, E.E. (Eds.), Readings in Cognitive Science. Elsevier, pp. 399–421. https://doi.org/10.1016/B978-1-4832-1446-7.50035-2.
Santangelo, M., Marchesini, I., Bucci, F., Cardinali, M., Fiorucci, F., Guzzetti, F., 2015. An approach to reduce mapping errors in the production of landslide inventory maps. Nat. Hazards Earth Syst. Sci. https://doi.org/10.5194/nhess-15-2111-2015.
Sekiguchi, T., Sato, H.P., 2006. Feature and distribution of landslides induced by the mid Niigata prefecture earthquake in 2004, Japan. J. Japan Landslide Soc. 43, 142–154. https://doi.org/10.3313/jls.43.142.
Shariati, M., Mafipour, M.S., Mehrabi, P., Bahadori, A., Zandi, Y., Salih, M.N.A., Nguyen, H., Dou, J., Song, X., Poi-Ngian, S., 2019. Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in Normal and high-strength concrete. Appl. Sci. 9, 5534. https://doi.org/10.3390/app9245534.
Shirzadi, A., Saro, L., Joo, O.H., Chapi, K., Hyun Joo, O., 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. Nat. Hazards https://doi.org/10.1007/s11069-012-0321-3.
Sidle, R.C., Bogaard, T.A., 2016. Dynamic earth system and ecological controls of rainfall-initiated landslides. Earth-Science Rev 159, 275–291. https://doi.org/10.1016/j.earscirev.2016.05.013.
Simon, N., Crozier, M., de Roiste, M., Rafek, A.G., 2013. Point based assessment: selecting the best way to represent landslide polygon as point frequency in landslide investigation. Electron. J. Geotech. Eng. 18 D, 775–784.
Steger, S., Brenning, A., Bell, R., Petschko, H., Glade, T., 2016. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. Geomorphology 262, 8–23.
Süzen, M.L., Doyuran, V., 2004. Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng. Geol. 71, 303–321. https://doi.org/10.1016/S0013-7952(03)00143-1.
Swingler, K., 1996. Applying Neural Networks: A Practical Guide. Morgan Kaufmann.
Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H., 2016. Generation of the 30 M-MESH global digital surface model by alos prism, in: international archives of the photogrammetry. Remote Sensing and Spatial Information Sciences-ISPRS Archives https://doi.org/10.5194/isprsarchives-XLI-B4-157-2016.
Tien Bui, D., Shirzadi, A., Shahabi, H., Geertsema, M., Omidvar, E., Clague, J.J., Thai Pham, B., Dou, J., Talebpour Asl, D., Bin Ahmad, B., Lee, S., 2019. New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. Forests 10, 743. https://doi.org/10.3390/f10090743.
Wang, F., Fan, X., Yunus, A.P., Siva Subramanian, S., Alonso-Rodriguez, A., Dai, L., Xu, Q., Huang, R., 2019. Coseismic landslides triggered by the 2018 Hokkaido, Japan (Mw 6.6), earthquake: spatial distribution, controlling factors, and possible failure mechanism. Landslides 16, 1551–1566. https://doi.org/10.1007/s10346-019-01187-7.
Yamagishi, H., Ito, Y., Kawamura, M., 2002. Characteristics of deep-seated landslides of Hokkaido; analyses of a database of landslides of Hokkaido, Japan. Environ. Eng. Geosci. 8 (1), 35–46. https://doi.org/10.2113/gseegeosci.8.1.35.
Yamagishi, H., Yamazaki, F., 2018. Landslides by the 2018 Hokkaido Iburi-Tobu earthquake on September 6. Landslides 15, 2521–2524. https://doi.org/10.1007/s10346-018-1092-z.
Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2015. 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. https://doi.org/10.1007/s10346-015-0614-1.
Yunus, A.P., Dou, J., Song, X., Avtar, R., 2019. Improved bathymetric mapping of coastal and Lake environments using Sentinel-2 and Landsat-8 images. Sensors 19, 2788. https://doi.org/10.3390/s19122788.
Yunus, A.P., Fan, X., Tang, X., Jie, D., Xu, Q., Huang, R., 2020. Decadal vegetation succession from MODIS reveals the spatio-temporal evolution of post-seismic landsliding after the 2008 Wenchuan earthquake. Remote Sens. Environ. 236, 111476. https://doi.org/10.1016/j.rse.2019.111476.