

Elsevier required licence: © 2020

This manuscript version is made available under the
CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The definitive publisher version is available online at

<https://doi.org/10.1016/j.catena.2019.104249>

Application of Convolutional Neural Networks Featuring Bayesian Optimization for Landslide Susceptibility Assessment

Maher Ibrahim Sameen¹, Biswajeet Pradhan^{1,2*}, Saro Lee^{3,4*}

¹Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, Building 11, Level 06, 81 Broadway, Ultimo, NSW 2007, PO Box 123, Australia

²Department of Energy and Mineral Resources EngineeringSejong University, Choongmu-gwan, 209 Neungdong-ro Gwangjin-gu, Seoul, 05006, South Korea

*Email. Biswajeet24@gmail.com or Biswajeet.Pradhan@uts.edu.au

³Geological Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), Gajeong-dong 30, Yuseong-gu, Daejeon, 305-350, South Korea

⁴Korea University of Science and Technology, 217 Gajeong-roYuseong-gu, Daejeon, 34113, Korea

*Email. leesaro@kigam.re.kr; Biswajeet.Pradhan@uts.edu.au

Abstract

This study developed a deep learning based technique for the assessment of landslide susceptibility through a one-dimensional convolutional network (1D-CNN) and Bayesian optimisation in Southern Yangyang Province, South Korea. A total of 219 slide inventories and 17 slide conditioning variables were obtained for modelling. The data showed a complex scenario. Some past slides have spread over steep lands, while others have spread through flat terrain. Random forest (RF) served to keep only important factors for further analysis as a pre-processing measure. To select CNN hyperparameters, Bayesian optimization was used. Three methods contributed to overcoming the overfitting issue owing to small training data in our research. The selection of key factors by RF helped first of all to reduce information dimensionality. Second, the CNN model with 1D convolutions was intended to considerably decrease the number of its parameters. Third, a high rate of drop-out (0.66) helped reduce the CNN parameters. Overall accuracy, area under the receiver operating characteristics curve (AUROC) and 5-fold cross-validation were used to evaluate the models. CNN performance was compared to ANN and SVM. CNN achieved the

highest accuracy on testing dataset (83.11%) and AUROC (0.880, 0.893, using testing and 5-fold CV, respectively). Bayesian optimization enhanced CNN accuracy by ~3% (compared with default configuration). CNN could outperform ANN and SVM owing to its complicated architecture and handling of spatial correlations through convolution and pooling operations. In complex situations where some variables make a non-linear contribution to the occurrence of landslides, the method suggested could thus help develop landslide susceptibility maps.

Keywords landslide susceptibility; GIS; deep learning; convolutional neural networks; Bayesian optimization

1. Introduction

A landslide is a form of mass wasting (generally downslope) related to the movement of natural materials (soil, rocks, or even debris; Leynaud et al., 2017). Unfortunately, landslides are unavoidable natural disasters. Landslides differ in size, but usually involve significant ground movements causing infrastructure damage (Chitu et al., 2016). They occur on- or off-shore; gravity drives debris and rocks movement to a landslide. Apart from gravity, other factors such as slope steepness, geological setting of the area, and human activities lead to landslides (Hadmoko et al., 2017).

Numerous research studies were conducted to identify mathematical or statistical links between the conditions of landslides and their probability of occurrence (Chen et al., 2019; Pham et al., 2019; Nsengiyumva et al., 2019). This kind of analysis is often known as susceptibility assessment/mapping or spatial modelling. Gökçeoglu and Aksoy (1996) investigated such analysis with fundamental image processing methods. Atkinson et al. (1998) has created linear modelling in central Apennines, Italy for spatial modelling of a landslide. For Derbyshire UK, Cross (1998) developed techniques to evaluate landslide susceptibility models. Since then, various new techniques and methods have been created to better model the susceptibility to landslides and to better understand the causes of landslides (Dehnavi et al., 2015; Hong et al., 2016; Youssef et al., 2016; Pourghasemi and Rossi, 2017; Reichenbach et al., 2018; Braun et al., 2019; Mohammady et al., 2012;). Overall, four primary kinds of landslide susceptibility models are developed including expert-based, statistical, machine learning, and a hybrid that combines two or more models from one or different categories (Fanous and Pradhan, 2019; Pradhan and Sameen, 2017; Pradhan et al.,

2017; Hong et al., 2017a, b; Tien Bui et al., 2016; Youssef et al., 2015). There are many models of each type. For more information on these models, the reader can refer to Pradhan et al. (2017). Various studies have shown that logistic regression (LR), support vector machines (SVM), and artificial neural networks (ANN) perform well in this context (Bui et al., 2016; Hong et al., 2016; Wang et al., 2016).

For the landslide susceptibility evaluation, several scientists have used ANN and a thorough assessment of this model can be found in Gorsevski et al., (2016). Other successful implementation of ANN-based models presented by Yesilnacar and Topal (2005) and Nefeslioglu et al. (2008). ANN was found superior to other models such as weights of evidence (WoE) (Wang et al., 2016), LR (Lee et al., 2016; Chen et al., 2016), Dempster-Shafer (Chen et al., 2016). ANN also found better than SVM, kernel LR, and a logistic model tree (LMT) in a study by Bui et al. (2016). In another geographic region, Yilmaz (2010) showed that ANN outperformed conditional probability, LR, and SVM. Although ANN considered a black-box model, it can provide the relative importance of the conditioning factors by several methods can be found in (Hines et al., 1997), (Lee et al., 2004), (Kanungo et al., 2006), and (Pradhan and Lee, 2010).

The above-mentioned studies implemented ANN models in landslide susceptibility assessment. However, recent developments of neural networks have shown that other model architectures, such as convolutional and recurrent, can significantly outperform the classical feedforward ANN. Convolutional neural networks or CNN was successfully applied to remote sensing, computer vision, voice recognition and natural linguistic processing (Maggiori et al., 2017; Sameen and Pradhan, 2017; LeCun et al., 2015). In the context of landslide research, CNN was used for landslide detection (Ding et al., 2016; Yu et al., 2017; Mezaal et al., 2017; Ghorbanzadeh et al., 2019). However, the performance of these models is not established for landslide susceptibility mapping (Wang et al., 2019). As a result, this research develops a model based on a one-dimensional CNN (1D-CNN) for landslide susceptibility mapping in north-east Korea, in the south of Yangyang Province. Also, comparing their performance to other standard feedforward ANN and SVM models.

2. Study Area and Data

Seoul is the capital and biggest town in South Korea. It has a large population density of over 10 million individuals. Natural hazards such as landslides are a significant threat in the region.

In this research, the region of concern is situated in Yangyang Province about 155 km west of Seoul. It includes about 53 km² (8.43 percent of Yangyang Province).

First, landslide inventory map was extracted from aerial photographs (<http://map.daum.net>). Additional data sources including previous landslide records and field research were also used to prepare landslide location data. Aerial photographs with a 0.5 m spatial resolution were obtained of Yangyang. A total of 219 landslides were identified through comparative analysis. The landslide data have been randomly split into training (70% of landslides; 164) and validation (30% of landslide; 55). The landslide inventory map is shown in Fig.1. Second, 17 landslide conditioning factors were derived from a number of sources (Table 1). These factors were in relation to natural processes in landslides and geo-environmental features of our research area. Some factors had categorical values (timber type, timber density, timber diameter, timber age, lithology, soil texture, land use) while others had continuous values. Those with continuous values were not reclassified to reduce model's sensitivities to reclassification methods. The geomorphological factors such as altitude, slope gradient, slope aspect, and curvature was derived from a 5 m digital terrain model created based on national topographic maps (scale, 1:5,000). Altitude affects the extent of rock weathering and considered a significant in landslide susceptibility analysis (Pradhan et al., 2014). The altitude map shows that the lowest and highest elevations in our study area were respectively -5,88 m and 929,79 m (Fig.2a). Most of the landslides occurred at greater altitudes and some at lower altitudes (in the area's northeast). Slope is another important factor and used almost in all studies related to susceptibility analysis (Aghdam et al., 2017). The slope factor was derived from the 5 m digital elevation model. The slope values ranged from 0 to 79° as shown in Fig. 2b. The slope aspect regulates topographic moisture concentrations influenced by both solar radiation and precipitation (Sadr et al., 2014). Fig.2c shows the aspect map; flat regions are stated by '-1' and other slope directions are in degrees from the north-east toward the north-west. In addition, three types of land curvatures namely total curvature, plan curvature, and profile curvatures were included in our analysis (Fig. 2d-f). Overall, curvature represents slope changes along a curve's tiny arcs, influencing slope instability. The curvature of the plane is the curvature perpendicular to the direction of the peak slope. The curvature is positive when the surface is convex and negative when concave. A zero-value shows linear surface. Instead, the profile curvature refers to the convergence and flow divergence across a surface. Combining plane and profile curvature is called total curvature.

Lithology and soil texture were extracted from the geology map provided by the Rural Development Administration (RDA). These factors have been used in many researches because they affect the type and mechanism of landslides as rocks differ in terms of internal structure, mineral composition, and susceptibility to landslides (Ercanoglu, 2005). In addition, human activities influence patterns of land use, contributing to landslides, more common in barren areas than forests and residential areas (Glade, 2003; Lallianthanga and Lalbiakmawia, 2013). Moreover, a polygon-based vegetation map (scale = 1: 25,000) was obtained from the Korean Forest Service. Four parameters were used as conditioning factors: timber type, timber density, timber diameter, and timber age (Fig. 5). Vegetation strengthens the soil by interlocking root networks forming erosion-resistant sloping mats.

The hydrological parameters assessed included the stream power index (SPI), the sediment transport index (STI), the topographic roughness index (TRI), and the topographic wetness index (TWI). The SPI is described as the motion of strong particles when gravity acts on sediments (Fig. 6a, Equation 1) (Jebur et al., 2014). The STI as shown in Fig.6b represents slope failure and deposition and is calculated using Equation 2 (Jebur et al., 2014). The TRI (Fig. 6c) defines the roughness of the local terrain and thus affects topographic and hydrological processes vital for landslide development. It influences landslide incidence and was calculated using the minimum and maximum cell values in the DTM's nine rectangular windows (Equation 3) (Jebur et al., 2014). TWI (Fig. 6d) measures topographic control of hydrological processes, reflecting slope and direction of flow (Equation 4) (Jebur et al., 2014).

$$SPI = A_s \times \tan\beta \quad (1)$$

$$STI = \left(\frac{A_s}{22.13}\right)^{0.6} \times \left(\frac{\sin\beta}{0.0896}\right)^{1.3} \quad (2)$$

$$TRI = \sqrt{\max^2 - \min^2}, \quad (3)$$

$$TWI = \ln\left(\frac{A_s}{\beta}\right) \quad (4)$$

where A_s is the specific catchment area (m^2/m), β is slope angle in degrees, max and min are the highest and minimum values of the cells in the nine rectangular windows of DTM. The specific

catchment area is upslope area per unit contour length, taken as the number of cells times grid cell size (cell area divided by cell size).

Fig. 1. Landslide inventory map of study area.

Table 1 List of landslide causal factors used in this study.

Fig. 2. Geomorphologic landslide conditioning factors, (a) altitude, (b) slope, (c) aspect, (d) total curvature, (e) plan curvature, and (f) profile curvature.

Fig. 3. Lithology (a) and soil (b) factors.

Fig. 4. Land use factor.

Fig. 5. Vegetation related factors, (a) timber type, (b) timber density, (c) timber diameter and (d) timber age

Fig. 6. Hydrological landslide conditioning factors, (a) SPI, (b) STI, (c) TRI, and (d) TWI.

3. Methods

Here we describe our landslide susceptibility model. The model featured an evaluation of relative importance of the conditioning factors with Random Forest (RF) (Pourghasemi and Rahmati, 2018; Sahin et al., 2018), modelling using 1D-CNN (Kussul et al., 2017; Zhu et al., 2017; Tang et al., 2019), and the model's architecture and hyperparameter optimization with a Bayesian-based approach (Sameen et al., 2018).

3.1 Preparing Datasets for Modelling

This landslide inventories were used for the production of non-landslide samples. The samples have been chosen to meet the following requirements. First, any non-landslide sample should be at least 500 meters away from landslides. Second, the distance between any two non-

landslide samples must be greater than 100 meters. These requirements help to prevent sampling in areas that have experienced landslide events. In total, 438 samples are then generated by merging the landslide and non-landslide sample points given that they encoded to 1 and 0, respectively.

In addition, the maps of landslide conditioning factors (17 of them) were prepared in GIS in raster format. Each factor was represented as a single band raster. Finally, our database included landslide conditioning factors as well as landslide/non-landslide samples. In order to apply 1D-CNN, multiple convolutional filters have been considered learning different portions of the data. The data were split from left to right across the layer from top to bottom, and a number of 1D sequences were generated. Landslide pixels were used to identify the 1D sequences that represent the landslide-prone locations. Instead, the non-landslide pixels were used to recognize the 1D sequences that represent the areas that are not prone to landslides. The data was then divided randomly into training (70%) and validation (30%). As a CNN input, the 1D sequences were used, which were generated using landslide conditioning factors. The 1D sequences generated by the sample's layers were transformed into one hot encoding and used as a CNN output. 1D-CNN computes a weighted sum of input layers, allowing to select certain combinations of features that are useful in next layers. For ANN and SVM, the values of the landslide conditioning factors were extracted at the point samples. Each sample is associated with its target label (1 for landslides, 0 for non-landslides).

3.2 Selection of Significant Conditioning Factors

A multicollinearity test did not lead to any elimination of variables. The relative importance of the landslide conditioning factors was evaluated using a conventional RF technique. RF is an ensemble method based on decision trees; RF produces (and then combines) many classification trees to calculate a classification by reference to the majority votes for individual decisions (Breiman, 2001). RF provides a good generalization, controlled by the number of trees used (Breiman, 2001). More importantly, RF estimates the relative importance of a predictive variable by identifying model errors in which the importance of the predictive variable under study is calculated. RF offers several benefits; features require no rescaling, transformation, or alteration; and less sensitive to outliers.

To select the best subset of the factors to include in susceptibility modelling, a percentage limit is required to select the subset proportion. Another alternative used in this study is to select various subset percentages and select the subset that provides the best accuracy using the validation dataset.

3.3 Modelling Algorithms

This research used different methods for landslide susceptibility assessment in the south of Yangyang including the proposed 1D-CNN, ANN, and SVM. These techniques are briefly described in this section.

3.3.1 Convolutional Neural Networks (1D-CNN)

CNN is a type of neural network that LeCun first created to analyse images and videos (LeCun et al., 1998). The main distinction between CNN and a traditional neural network is that former benefits from the properties of natural signals such as local connections, shared weights, pooling, and the use of multiple layers (LeCun et al., 2015). The model is constructed in a series of layers, the very first being convolutional and pooling layers. In the convolutional layers, units are structured into feature maps and connected via filter banks to local patches in the feature map of the previous layer. The output of this process is then passed through ReLU – a non-linearity activation function. In this way, CNN identifies information that is correlated and invariant to location within local groups of values. On the other hand, pooling layers often compute peak values of local unit patches in feature maps and are used to combine features that are semantically similar. Several stages of convolution, nonlinearity, and pooling are stacked, followed by a few fully connected layers and finally a softmax layer that engages in classification or prediction. CNN (like other neural networks) is generally trained via backpropagation and stochastic gradient descent algorithms.

3.3.2 Feed forward Artificial Neural Networks (ANN)

ANN is a model inspired by how biological nervous systems (e.g., the brain) operate. It consists of several interconnected neurons working together to solve specific problems. ANN's main idea is the learning process. Learning in biological systems includes adjusting synaptic neuronal links. That is also true of an ANN. Stacking layers build an ANN; each layer mimics an

array of neurons. A typical example is the feedforward multilayer perceptron (MLP). MLP networks usually have three layers of processing elements and only one hidden layer but adding more hidden layers to the network is not restricted. The input layer's task is to receive external stimuli and propagate them to the next layer. Hidden layers get weighted sums of incoming signals and process them using an activation function, most commonly saturation, sigmoid, and hyperbolic tangent functions. In turn, hidden units send output signals to the next layer. This can either be another hidden layer or the output layer. Information is propagated forward until the network produces output.

3.3.3 Support Vector Machines (SVM)

SVM is a supervised model based on the minimization of structural risk (Wan et al., 2010; Yao et al., 2008). SVM features a high-dimensional hyperplane or set of hyperplanes and can be used for classification and regression. It classifies training data into two categories. SVMs can perform non-linear classification efficiently using the “kernel trick”, implicitly mapping inputs into high-dimensional feature spaces (Table 2).

Table 2 The common kernel functions used with SVM.

Consider a training dataset (x_i, y_i) with $x_i \in R^n$, $y_i \in \{1, -1\}$, and $i = 1, \dots, m$. When the data are linearly separable, a separating hyper-plane can be defined as

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i, \quad (5)$$

where w is a coefficient vector that defines the orientation of the hyperplane in the feature space, b is the offset of the hyperplane from the origin, and ξ_i is the positive slack variables (Cortes and Vapnik, 1995). The following optimization problem using Lagrangian multipliers is solved by determining the optimal hyper-plane (Samui, 2008).

$$\text{Minimize } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i x_j), \quad (6)$$

$$\text{Subject to } \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad (7)$$

where α_i are Lagrange multipliers, C is the penalty, and the slack variables ξ_i allows for penalized constraint violation. The decision function used for classification of new data can be written as

$$g(x) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_i x_i + b \right) \quad (8)$$

If separation of the hyperplane using the linear kernel function is not possible, the original input data may be shifted into a high-dimensional feature space using certain nonlinear kernel functions. The classification decision function is then written as follows:

$$g(x) = \text{sign} \left(\sum_{i=1}^n y_i \alpha_j K(x_i, x_j) + b \right) \quad (9)$$

where $K(x_i, x_j)$ is the kernel function.

3.4 Bayesian Optimization

Deep learning techniques like CNN involve many hyperparameters optimization including batch size, activation function, an optimization algorithm, and network structure (depth and a number of hidden layers). Since most hyperparameters are continuous variables with only loose constraints on their numerical ranges, selecting them is a basic challenge in deep learning research. Trial-and-error optimization approaches are slow; hyperparameters have coupling effects, and the optimal ranges cannot be quickly identified. Robust optimization techniques are needed.

Three techniques for optimizing hyperparameters are frequently used to fine-tune machine-learning algorithms; these are grid search, random search, and Bayesian optimization. The grid search method optimizes all possible parameter settings to select the best combination of algorithm parameters. This is time-consuming when the number of hyperparameters and search spaces are comparatively large (as with neural networks). Furthermore, the operator must define particular parameter search space values. The random search technique selects random hyperparameter values, then evaluates model accuracy. Evaluations run iteratively; the operator selects the number

of runs. After evaluating all runs, using an assessment metric, the best hyperparameter setup is chosen. In this technique (as in grid search), there is no guarantee that the next assessment will be better than the prior setup. The significant issue with grid and random search optimization techniques is ignoring historical information.

Bayesian optimization, as the word suggests, optimizes decision-making regarding which parameter setup to assess next (Sameen et al., 2018). In this context, an objective function is used to know how prior settings (exploitation) were conducted. Bayesian optimization techniques find the best possible parameter setup faster than grid and random searches. Training of deep learning algorithms (e.g., CNN) can be highly time-consuming considering the quantity of data concerned and the computational density needed. In such circumstances, Bayesian optimization can be particularly helpful. We used Bayesian optimization to find a sub-optimal ANN, SVM, and CNN hyperparameters. Table 3 summarizes these hyperparameters and their search spaces.

The Bayesian optimization in our study was based on 25 iterations, also known as a number of calls (n). Negative minimum AUROC with 5-fold cross validation was used as objective function. After each iteration, better model's configuration was found until convergence at the 25 call. We did not see significant improvements in more iterations. The parameters found at the iteration 25 thus considered best in our case study. At the cost of extra computing time, one may find better configurations if searches for larger search spaces.

Table 3 Hyperparameters and their search space of the proposed models.

3.5 Evaluation Methods

Many performance evaluation forms can be used to evaluate landslide susceptibility models. Three such metrics were used: overall accuracy (OA) (Hong et al., 2018), the area under the receiver operating characteristic curve (AUROC), and k -fold cross-validation. OA was calculated based on the number of correctly classified landslides in terms of the total numbers of landslides used in training/validation. ROC curves plot sensitivity on y-axis and specificity on x-axis, corresponding to decision thresholds (Fawcett, 2006; Dou et al., 2014). Sensitivity is the probability that landslide pixels are classified correctly as 'landslide'. Specificity is the probability that non-landslide pixels are classified correctly as 'non-landslide'. AUROC is the area under the ROC curve, indicating the goodness-of-fit of a landslide model (Jones and Athanasiou, 2005). A

perfect model has an AUROC of 1 and an AUROC of 0.5 indicates random performance. The closer the AUROC to 1, the better the model performance (Walter, 2002; Pourghasemi et al., 2012).

4. Results

We present and discuss our main outcomes here. All experiments were implemented in Python using Scikit-Learn and Keras on a computer with a Core i7-4510U CPU running at 2.60 GHz, 16 GB of RAM, and a x64-based processor.

4.1 Selecting Significant Conditioning Factors

This study initially used 17 factors, but there was no assurance that all factors are significant. Multicollinearity test and feature importance were analyzed. The multicollinearity test did not lead to any elimination of variables. We studied the feature importance using RF. The number of trees was initially set at 500 but converged to 130. After training, for each feature, the relative importance and rank were calculated. Table 4 summarizes the outcomes, importance values, and ranking for the landslide conditioning factors. The three most significant factors were altitude (importance = 1.000), soil (0.869), and plan curvature (0.815). The three minor factors were aspect (0.530), timber density and timber diameter (0.269, 0.353). In addition, the best subset percentage (12/17) was selected based on the average validation accuracy of all the models. It yielded 75.89% and served as susceptibility model inputs.

Table 4 Random Forest importance values and ranks estimated for the landslide conditioning factors.

4.2 Model Performance

CNN performance was assessed over 500 epochs with early termination enabled, monitoring validation accuracy using 0.01 *min_delta* and 20 *patience*. The *min_delta* is the minimum change in the validation accuracy that counts as an improvement (i.e., an absolute change of less than the *min_delta* showed no improvement). *Patience* is the number of epochs exhibiting no improvement after which training ceased. Early termination mode was set to “auto” to allow automatic reduction/increase in validation accuracy.

Fig. 7 shows model losses and accuracies over 63 epochs using training and validation datasets. Training was finished after 63 epochs due to no further improvement in validation accuracy. Model loss with both training and validation datasets reduced over time, suggesting that the model was learning from the data. In other words, after several epochs, CNN learned how to predict landslides using the input factors. Similarly, model training and validation accuracies increased over time as learning performance improved. Fluctuations in model loss and accuracy over distinct epochs are attributable to hidden layers dropout. In addition, the low gap between training and validation curves suggests that overfitting was minimal; dropout was used to avoid overfitting and to increase computational performance. The (minimum, maximum) model loss and accuracy using the validation dataset were (0.477, 0.585) and (59.20%, 84.30%), respectively.

Fig. 7. Performance of the CNN model.

4.3 Results of Optimization

Model hyperparameters was chosen via Bayesian optimization, which after fewer evaluations produced good settings with no expert intervention. Table 5 summarizes the best configurations for the different models. The optimization was run 25 times to find the best parameter values with specific search spaces. No significant improvements were apparent in extra iterations. Configuration always improved after each iteration. After 25 iterations, the best model setup was found.

As shown in Fig. 8, Bayesian optimization achieved good results after a few iterations. The lines in the figure are the values of an objective function (i.e., validation accuracy) acquired by the best-performing setup after completing all evaluations. ANN's accuracy improved from 72.09% to 76.20%. In addition, from the initial setup, CNN's accuracy grew by ~3%.

Fig. 8. The best-found configurations for ANN and CNN by Bayesian optimization.

Table 5 The best-selected hyperparameter values of different models.

4.4 Sensitivity Analysis

This sub-section describes the impacts of model hyperparameters (ANN and CNN) on the accuracy of landslide susceptibility modelling. These included the optimization algorithm (or

optimizer), *activation function* used by hidden layers, a number of neurons in hidden layers, batch size, and dropout rate. CNN had two extra parameters: number of convolutional filters and the length of the input sequence.

Fig. 9 shows the impact of the *optimizer* and *activation function* on the accuracy. For both ANN and CNN, the optimization algorithm and the activation type considerably impacted model accuracy on validation dataset. In our analysis, the remaining parameters were kept at their optimal values. The best *optimizers* for models were “Adamax” and “Adagrad,” respectively. The lowest accuracy associated with SGD. Other *optimizers* provided similar results for both ANN and CNN. The best activation function was “ReLU” for ANN and “ELU” for CNN.

Fig. 10 shows the impacts of hidden layer neuron numbers and batch size used to train ANN and CNN. In the ANN hidden layers, the best neuron numbers were 311, and 240 in CNN. For ANN, the use of only four neurons in hidden layers provided the worse result (< 60%). However, the model performed well using more than 16 hidden neurons. In contrast, CNN performed likewise with all tested figures of hidden neurons. The lowest and highest accuracies were 75.15% and 83.11%, respectively. Best batch size values for ANN and CNN were 71 and 117, respectively. ANN was more sensitive to batch size values given its higher value of standard deviation (1.46) compared to that of CNN (0.83). Furthermore, the analysis showed that dropout rates of 0.57 and 0.66 were optimal for ANN and CNN.

CNN featured extra parameters (sequence length and a number of filters) (Fig. 11). The same optimization approach was used to find the best values of these parameters. The best sequence length was 10 (83.11%) compared to that of 3 (80.83%). CNN was much more sensitive to filter amount than sequence length.

Fig.9. The effects of optimizer and activation function on the performance of ANN and CNN.

Fig. 10. The effects of the number of neurons in the hidden layers, batch size and dropout rate on the performance of ANN and CNN.

Fig. 11. The effects of the sequence length and the number of filters on CNN performance.

4.5 Model Comparisons

CNN model performance was compared to other benchmark techniques, including ANN and SVM. The landslide susceptibility maps using ANN, SVM and CNN are shown in Fig. 12. Table 5 shows the model performance using training and validation datasets. NN-based models including ANN and CNN did better than SVM. CNN achieved the highest training (83.50%) and validation (83.11%) accuracy. ANN's training accuracy was 82.60%. Model performance was also investigated by constructing ROCs and calculating AUROC, and 5-fold cross-validation. These findings verified CNN's highest performance with both validation and cross-validation datasets. CNN's training, validation, and cross validation AUROC were 0.881, 0.88, and 0.893, respectively. SVM achieved AUROC 0.858 and 0.808 on training and validation datasets.

Table 6 Performance of the proposed deep learning models and the benchmark methods using the training and validation datasets.

4.6 Computational Performance

Table 7 summarizes the times (in seconds) required for the model's optimization, training, and making predictions on new data. Optimization was most time-consuming. It took 3,408.83 to optimize the CNN model. SVM optimization only required 9.27 s. In terms of training time, CNN and ANN took 0.59 s and 0.22 s, respectively, and only 0.08 s of SVM. However, the times required to make predictions based on new data were relatively short. If training is big, graphical processing units (GPUs) or cloud computing services such as Amazon AWS and Google Cloud among others may be required.

Fig. 12. Landslide susceptibility maps of the study area (a) ANN, (b) SVM, and (c) CNN.

Table 7 Computing performance of the models in seconds.

5. Conclusion

Deep learning has got a lot of attention from many scientists, including remote sensing and GIS. Deep learning was effectively implemented in several remote sensing applications such as classification, object detection, and feature selection. This research used CNN to create a landslide map for the South Yangyang Province. Since CNN includes many hyperparameters to be tuned, Bayesian optimization was used to search for their best values. Model performance was compared to ANN and SVM.

Yangyang region data showed a complicated scenario. Some previous landslide occurrences were spread in high-steep lands, while others in low-steep lands. To model such information, a comparatively complicated model is required to identify the non-linear relationship between landslide conditioning variables and the study area landslide occurrence. Our findings showed that CNN could outperform ANN and SVM owing to its complicated architecture and handling of landslide spatial correlations through convolution and pooling operations. This is also a sign of why SVM results are not very accurate or at least less accurate than CNN. SVM is comparable to a one-hidden layer neural network, which is obviously not complex enough to model such a scenario.

However, when models get complicated, the problem of overfitting arises. To handle this issue in our study, CNN was designed with 1D convolutions to reduce the number of its parameters significantly. We also used a high dropout rate (0.66) which contributed in reducing the number of CNN parameters. These techniques were found useful in dealing with small datasets as we had in our study. Not forgetting that selecting importance factor (12 in our case) could also reduce the dimensionality of the data, thus playing a significant role in avoiding model overfitting due to small training samples.

CNN is a promising solution to evaluating landslide susceptibility. Non-significant conditioning factors are suggested to be removed prior CNN application. We found no multicollinearity problem, but RF revealed that five conditioning variables (vegetation-related and aspect factors) were not important statistically; excluded from further analysis. After 25 iterations, Bayesian optimization recognized sub-optimal CNN hyperparameters in 3,408.83s, enhancing precision by 3%. Sensitivity analysis showed that some parameters (such as filter numbers in convolutional layers) and sequence length had greater impacts on CNN accuracy than other parameters.

A comparative research showed that CNN had the best OA, AUROC (validation and cross-validated datasets) outcomes relative to ANN and SVM. Overall, CNN performed well considering the assessment metrics used, and extra training and optimization could further boost efficiency. Our findings help investigate other deep learning models (recurring neural networks, generative models, and auto encoders) in terms of assessing landslide susceptibility. The landslide community must have these fresh models accessible. Our findings open several research directions for using synthetic data and sophisticated algorithms, as well as herald novel apps.

Acknowledgments

This research was conducted by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) funded by the Ministry of Science, ICT. This research (NRF-2016K1A3A1A09915721) was supported by Science and Technology Internationalization Project through National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT. Also, this research is supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), UTS, Australia.

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:

<http://www.textcheck.com/certificate/FnpUNa>

References

- Aghdam, I.N., Pradhan, B., Panahi, M., 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. *Environmental Earth Sciences*. 76(6), 237.
- Alin, A., 2010. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2(3), 370-374.
- Atkinson, P. M., & Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Computers & Geosciences*, 24(4), 373-385.

- Braun, A., Urquia, E. L. G., Lopez, R. M., & Yamagishi, H. (2019). Landslide Susceptibility Mapping in Tegucigalpa, Honduras, Using Data Mining Methods. In IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018-Volume 1 (pp. 207-215). Springer, Cham.
- Breiman, L., 2001. Random forests. *Machine learning*. 45(1), 5-32.
- Bui, D.T., Tuan, T.A., Klempe, H., Pradhan, B., Revhaug, I., 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(2), 361-378.
- Chen, W., Panahi, M., Tsangaratos, P., Shahabi, H., Ilia, I., Panahi, S., ... & Ahmad, B. B. (2019). Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *Catena*, 172, 212-231.
- Chen, W., Pourghasemi, H.R., Kornejady, A., Zhang, N., 2017. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma*. 305, 314-327.
- Chen, W., Pourghasemi, H.R., Zhao, Z., 2016. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto International*. 1-19.
- Chen, Z., Gao, B., Devereux, B., 2017 State-of-the-Art: DTM Generation Using Airborne LIDAR Data. *Sensors*. 17(1), 150.
- Chitu, Z., Busuioc, A., Burcea, S., Sandric, I., 2016. Landslide triggering rainfall thresholds estimation using hydrological modelling of catchments in the Ialomita Subcarpathians, Romania. In EGU General Assembly Conference Abstracts (Vol. 18, p. 12497).
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning*. 20, 273-297.
- Cross, M. (1998). Landslide susceptibility mapping using the matrix assessment approach: a Derbyshire case study. Geological Society, London, Engineering Geology Special Publications, 15(1), 247-261.
- Dehnavi, A., Aghdam, I.N., Pradhan, B., Varzandeh, M.H.M., 2015. A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *Catena*. 135, 2015, 122-148. <https://doi.org/10.1016/j.catena.2015.07.020>

- Ding, A., Zhang, Q., Zhou, X., Dai, B., 2016. Automatic recognition of landslide based on CNN and texture change detection. In Chinese Association of Automation (YAC), Youth Academic Annual Conference of (pp. 444-448). IEEE.
- Ercanoglu, M. (2005). Landslide susceptibility assessment of SE Bartın (West Black Sea region, Turkey) by artificial neural networks. *Natural Hazards and Earth System Sciences*, 5(6), 979-992.
- Fanos, A.M., Pradhan, B., 2019. A Spatial Ensemble Model for Rockfall Source Identification From High Resolution LiDAR Data and GIS. *IEEE Access*. 7, pp. 74570 – 74585. Doi: 10.1109/ACCESS.2019.2919977
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*. 102(3), 99-111.
- Fernández, C. I., Del Castillo, T. F., Hamdouni, R. E., & Montero, J. C. (1999). Verification of landslide susceptibility mapping: a case study. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 24(6), 537-544.
- Ghorbanzadeh, O., Hölbling, D., Meena, S. R., & Blaschke, T. (2019). Detecting Earthquake-triggered Large-scale Landslides with Different Input Window Sizes Convolutional Neural Networks.
- Glade, T. (2003). Landslide occurrence as a response to land use change: a review of evidence from New Zealand. *Catena*, 51(3-4), 297-314.
- Gökceoglu, C., & Aksoy, H. (1996). Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Engineering Geology*, 44(1-4), 147-161.
- Gorsevski, P.V., Brown, M.K., Panter, K., Onasch, C.M., Simic, A., Snyder, J., 2016. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: a case study in the Cuyahoga Valley National Park, Ohio. *Landslides*. 13(3), 467-484.
- Hadmoko, D.S., Lavigne, F., Samodra, G., 2017. Application of a semiquantitative and GIS-based statistical model to landslide susceptibility zonation in Kayangan Catchment, Java, Indonesia. *Natural Hazards*. 87(1), 437-468.
- Han, J., Pei, J., Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.

- Hines, J., Tsoukalas, L.H., Uhrig, R.E., 1997. MATLAB supplement to fuzzy and neural approaches in engineering. John Wiley & Sons, Inc.
- Hong, H., Pourghasemi, H. R., & Pourtaghi, Z. S. (2016). Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology*, 259, 105-118.
- Hong, H., Pradhan, B., Jebur, M.N., Bui, D.T., Xu, C., Akgun, A., 2016. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environmental Earth Sciences*. 75(1), 40.
- Hong, H., Pradhan, B., Sameen, M.I., Kalantar, B., Zhu, A., Chen, W., 2017a. Improving the accuracy of landslide susceptibility model using a novel region-partitioning approach. *Landslides*. 1-20.
- Hong, H., Liu, J., Zhu, AX., Shahabi, H., Pham, B.T., Chen, W., Pradhan, B., Tien Bui, D., 2017b. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China), *Environmental Earth Sciences*. 76: 652. <https://doi.org/10.1007/s12665-017-6981-2>
- Jebur, M. N., Pradhan, B., & Tehrany, M. S. (2014). Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment*, 152, 150-165.
- Kanungo, D.P., Arora, M.K., Sarkar, S., Gupta, R.P., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*. 85(3), 347-366.
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778-782.
- Lallianthanga, R. K., Lalbiakmawia, F., & Lalramchuana, F. (2013). Landslide hazard zonation of Mamit town, Mizoram, India using remote sensing and GIS techniques. *International Journal of Geology, Earth and Environmental Sciences*, 3(1), 184-194.
- Le, L., Lin, Q., Wang, Y., 2017. Landslide susceptibility mapping on a global scale using the method of logistic regression. *Natural Hazards and Earth System Sciences*. 17(8), 1411.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*. 521(7553), 436-444.

- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Aghdam, I.N., Pradhan, B., Panahi, M., 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. *Environmental Earth Sciences*. 76(6), 237.
- Lee, J.H., Sameen, M.I., Pradhan, B., Park, H.J., 2017. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology*.
- Lee, S., Jeon, S.W., Oh, K.Y., Lee, M.J., 2016. The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea. *Open Geosciences*. 8(1), 117-132.
- Lee, S., Ryu, J.H., Won, J.S., Park, H.J., 2004. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Engineering Geology*. 71(3), 289-302.
- Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., Moreau, J., 1995. Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*. 46(8), 1229-1236.
- Leynaud, D., Mulder, T., Hanquiez, V., Gonthier, E., Régert, A., 2017. Sediment failure types, preconditions and triggering factors in the Gulf of Cadiz. *Landslides*. 14(1), 233-248.
- Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*. 55(2), 645-657.
- Mezaal, M.R., Pradhan, B., Sameen, M.I., MohdShafri, H.Z., Yusoff, Z.M., 2017. Optimized Neural Architecture for Automatic Landslide Detection from High Resolution Airborne Laser Scanning Data. *Applied Sciences*. 7(7), 730.
- Mohammady, M., Pourghasemi, H.R., Pradhan, B., 2012. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. *Journal of Asian Earth Sciences*. 61(15), 221-236. <https://doi.org/10.1016/j.jseae.2012.10.005>
- Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., Ghazali, A.H., 2017. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-

- sensing data and GIS, *Geomatics, Nat. Hazards and Risk*. 8:2, 1080-1102, DOI: 10.1080/19475705.2017.1294113
- Naghibi, S.A., Moghaddam, D.D., Kalantar, B., Pradhan, B., Kisi, O., 2017. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. *Journal of Hydrology*. 548, 471-483. <https://doi.org/10.1016/j.jhydrol.2017.03.020>
- Nefeslioglu, H.A., 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. *Engineering Geology*. 97(3), 171-191.
- Nsengiyumva, J. B., Luo, G., Amanambu, A. C., Mind'je, R., Habiyaremye, G., Karamage, F., ... & Mupenzi, C. (2019). Comparing probabilistic and statistical methods in landslide susceptibility modeling in Rwanda/Centre-Eastern Africa. *Science of the Total Environment*, 659, 1457-1472.
- Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*. 154(1), 135-150.
- Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H., & Bui, D. T. (2019). Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena*, 175, 203-218.
- Pham, B.T., Bui, D.T., Dholakia, M.B., Prakash, I., Pham, H.V., Mehmood, K., Le, H.Q., 2017. A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomatics, Natural Hazards and Risk*. 8(2), 649-671.
- Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M.B., 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, B.T., Pradhan, B., Bui, D.T., Prakash, I., Dholakia, M.B., 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India). *Environmental Modelling & Software*. 84, 240-250.
- Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: which algorithm, which precision?. *Catena*, 162, 177-192.

- Pourghasemi, H. R., & Rossi, M. (2017). Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theoretical and Applied Climatology*, 130(1-2), 609-633.
- Pradhan, A. M. S., & Kim, Y. T. (2014). Relative effect method of landslide susceptibility zonation in weathered granite soil: a case study in Deokjeok-ri Creek, South Korea. *Natural hazards*, 72(2), 1189-1217.
- 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. *Environmental Modelling & Software*. 25(6), 747-759.
- Pradhan, B., Lee, S., 2010. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides*. 7(1), 13-30.
- Pradhan, B., Sameen, M.I., 2017. Landslide Susceptibility Modeling: Optimization and Factor Effect Analysis. In *Laser Scanning Applications in Landslide Assessment* (pp. 115-132). Springer International Publishing.
- Pradhan, B., Seenii, M.I., Kalantar, B., 2017. Performance Evaluation and Sensitivity Analysis of Expert-Based, Statistical, Machine Learning, and Hybrid Models for Producing Landslide Susceptibility Maps. In *Laser Scanning Applications in Landslide Assessment* (pp. 193-232). Springer International Publishing.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based statistically based landslide susceptibility models. *Earth-science reviews*, 180, 60-91.
- Sahin, E. K., Colkesen, I., & Kavzoglu, T. (2018). A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto International*, 1-23.
- Sameen, M. I., Pradhan, B., & Lee, S. (2018). Self-learning random forests model for mapping groundwater yield in data-scarce areas. *Natural Resources Research*, 1-19.
- Sameen, M.I., Pradhan, B., 2017. Severity Prediction of Traffic Accidents with Recurrent Neural Networks. *Applied Sciences*. 7(6), 476.
- Samui, P., 2008. Slope stability analysis: a support vector machine approach. *Environmental Geology*. 56, 255-267.

- Tang, H., Li, Y., Han, X., Huang, Q., & Xie, W. (2019). A Spatial-Spectral Prototypical Network for Hyperspectral Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*.
- Tien Bui, D., Ho, T.C., Pradhan, B., Pham, B-T., Nhu, V-H., Revhaug, I., 2016. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environmental Earth Sciences*. 75(14), 1102. <https://doi.org/10.1007/s12665-016-5919-4>
- Wan, S., Lei, T., Chou, T.Y., 2010. A novel data mining technique of analysis and classification for landslide problems. *Nat. Hazards*. 52, 211–230.
- 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. *Geosciences Journal*. 20(1), 117-136.
- Wang, Q., Li, W., Xing, M., Wu, Y., Pei, Y., Yang, D., Bai, H., 2016. Landslide susceptibility mapping at Gongliu county, China using artificial neural network and weight of evidence models. *Geosciences Journal*. 20(5), 705-718.
- Wang, Y., Fang, Z., & Hong, H. (2019). Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Science of The Total Environment*, 666, 975-993.
- Yao, X., Tham, L., Dai, F., 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology*. 101, 572-582.
- Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*. 79(3), 251-266.
- Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environmental Earth Sciences*. 61(4), 821-836.
- Youssef, A.M., Al-Kathery, M., Pradhan, B., 2015. Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. *Geosciences Journal*. 19(1), 113-134. <https://doi.org/10.1007/s12303-014-0032-8>
- Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (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(5), 839-856.
- Yu, H., Ma, Y., Wang, L., Zhai, Y., Wang, X., 2017. A landslide intelligent detection method based on CNN and RSG_R. In *Mechatronics and Automation (ICMA), 2017 IEEE International Conference on* (pp. 40-44). IEEE.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.

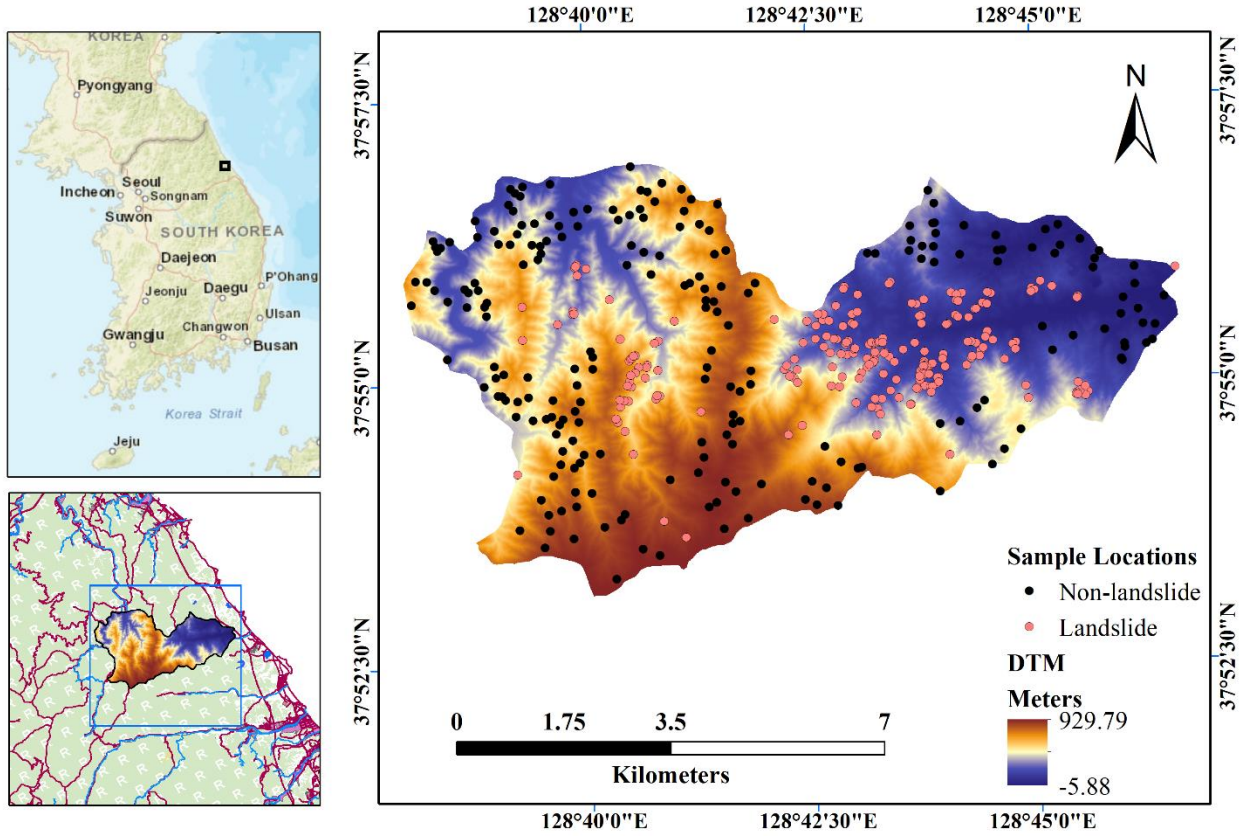


Fig. 1 Landslide inventory map of study area.

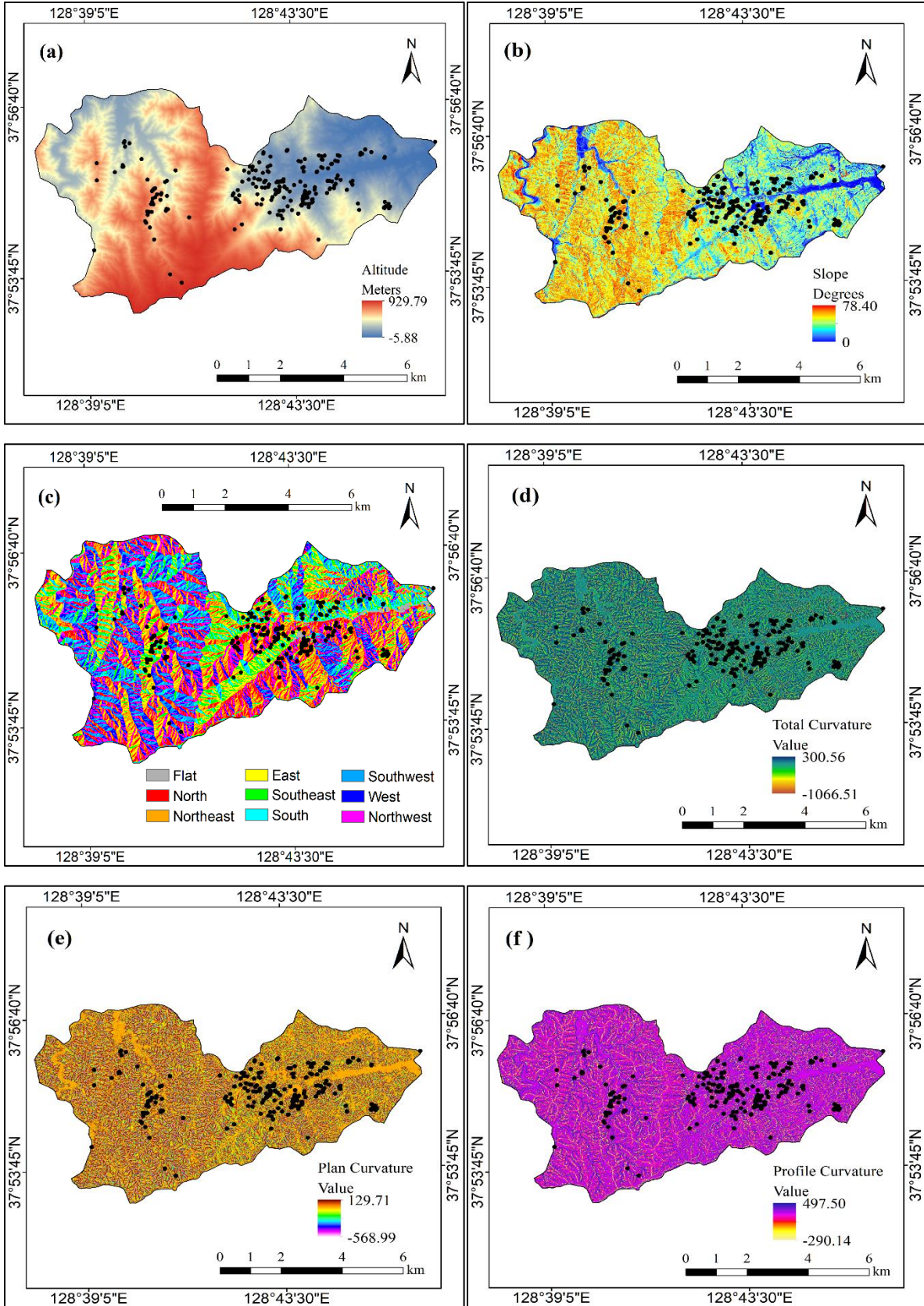


Fig. 2 Geomorphologic landslide conditioning factors, (a) altitude, (b) slope, (c) aspect, (d) total curvature, (e) plan curvature, and (f) profile curvature.

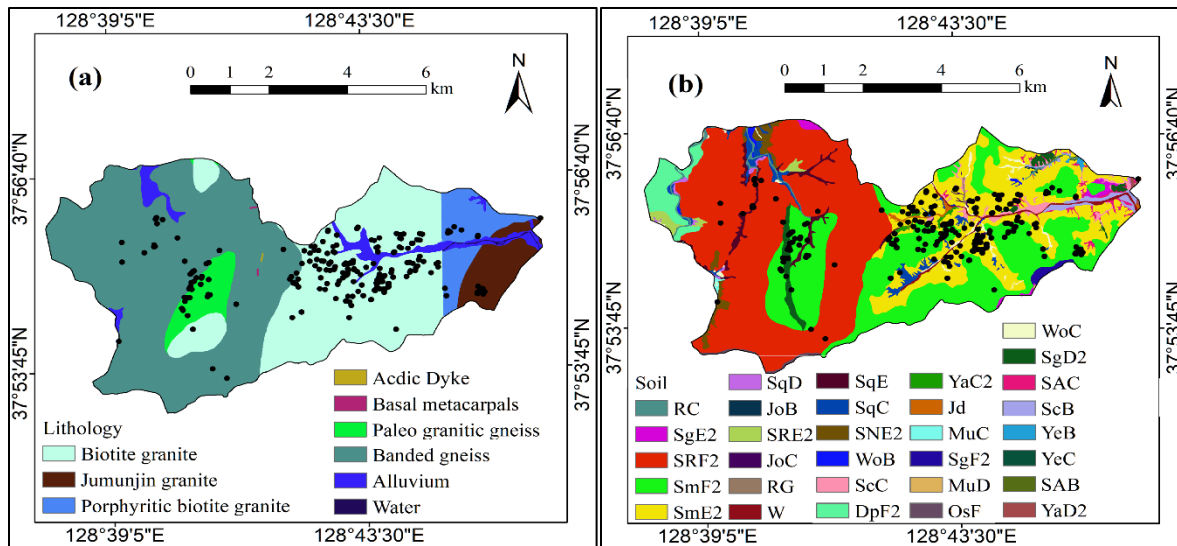


Fig. 3 Lithology (a) and soil (b) factors.

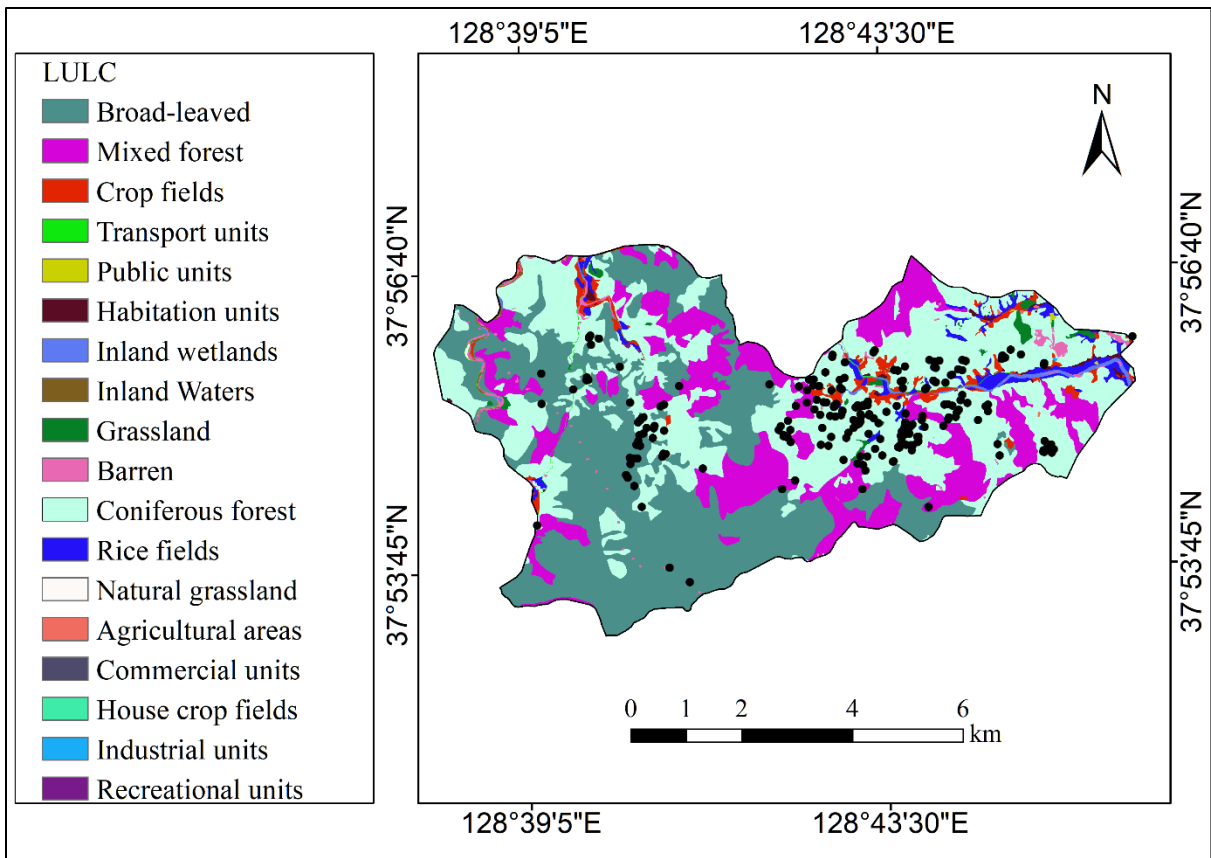


Fig. 4 Land use factor.

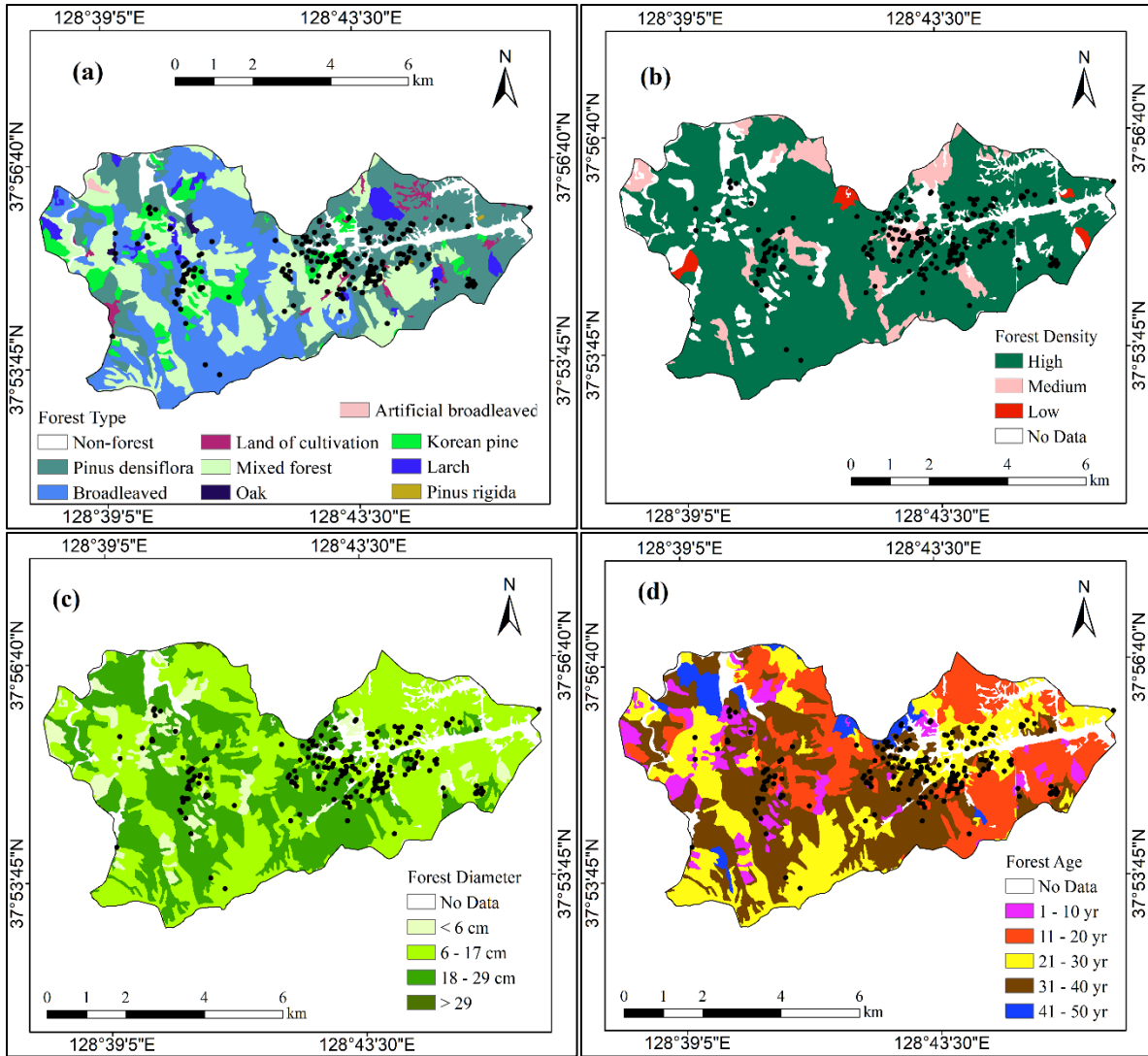


Fig. 5 Forest-related landslide conditioning factors, (a) forest type, (b) forest density, (c) forest diameter, and (d) forest age.

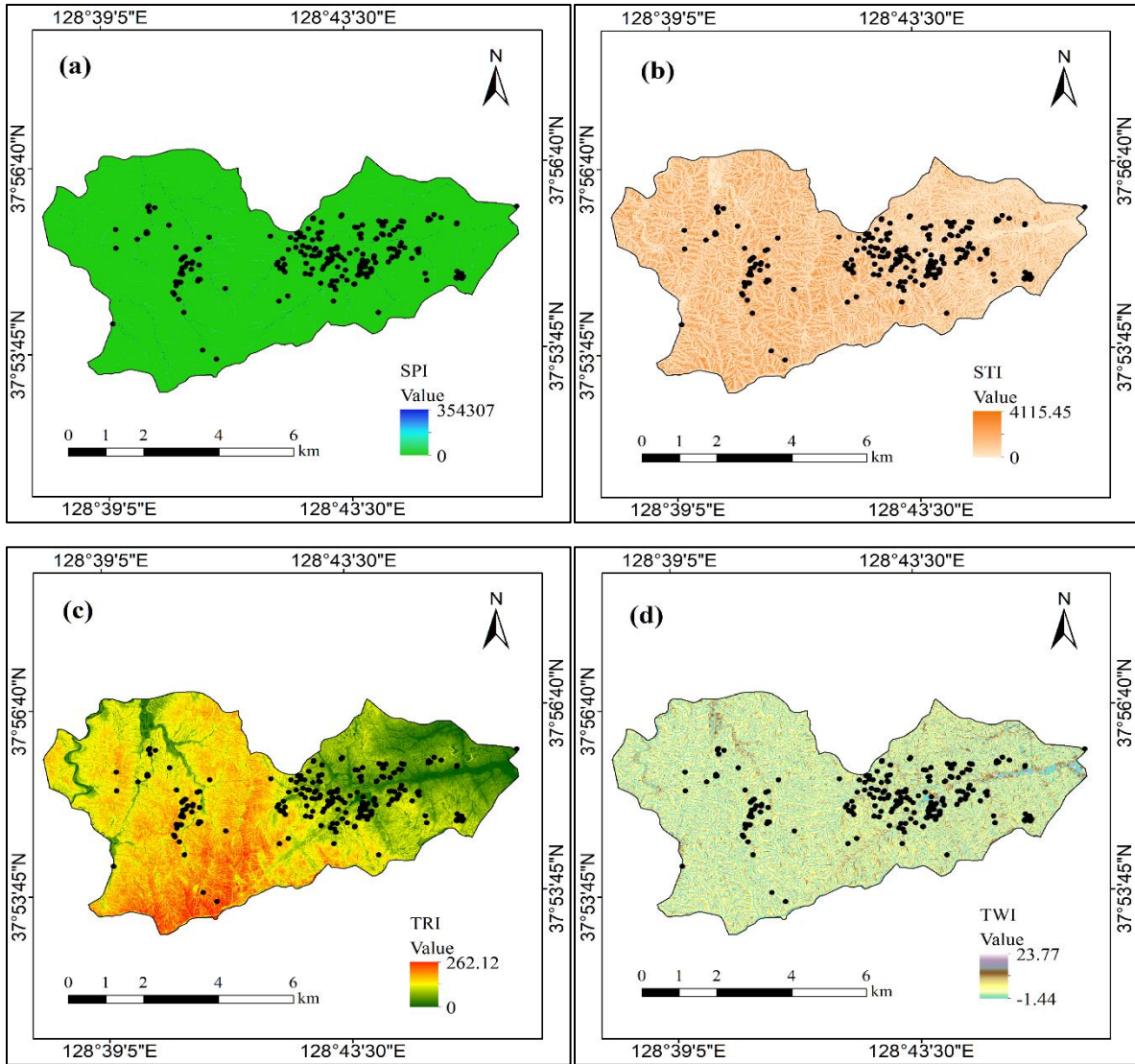


Fig. 6 Hydrological landslide conditioning factors, (a) SPI, (b) STI, (c) TRI, and (d) TWI.

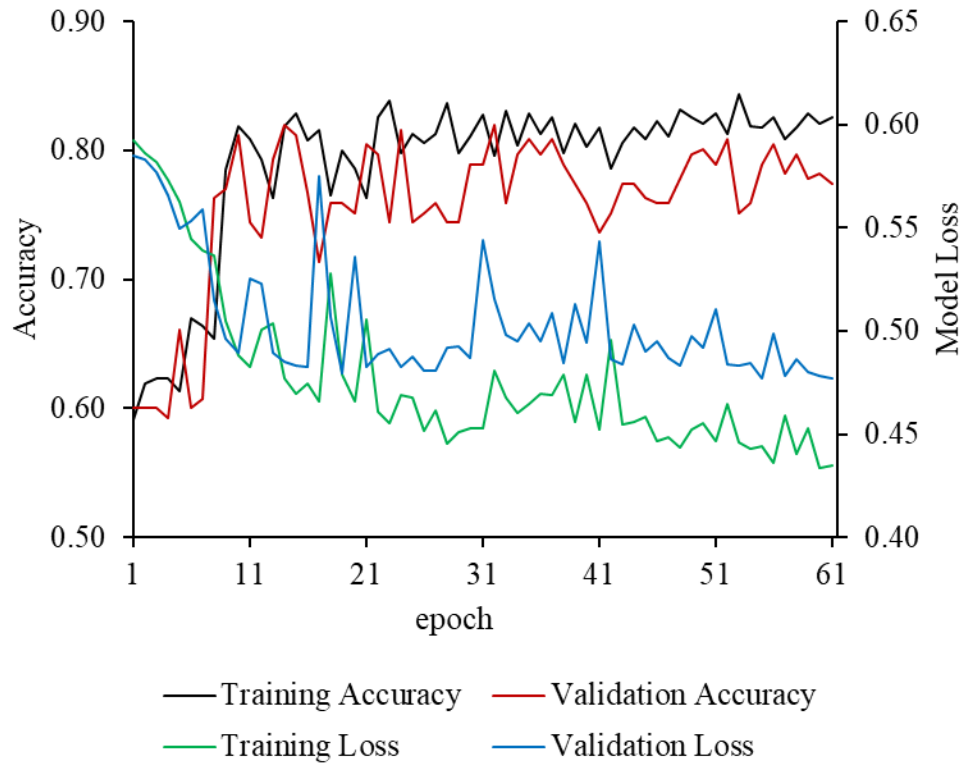


Fig. 7 Performance of the CNN model.

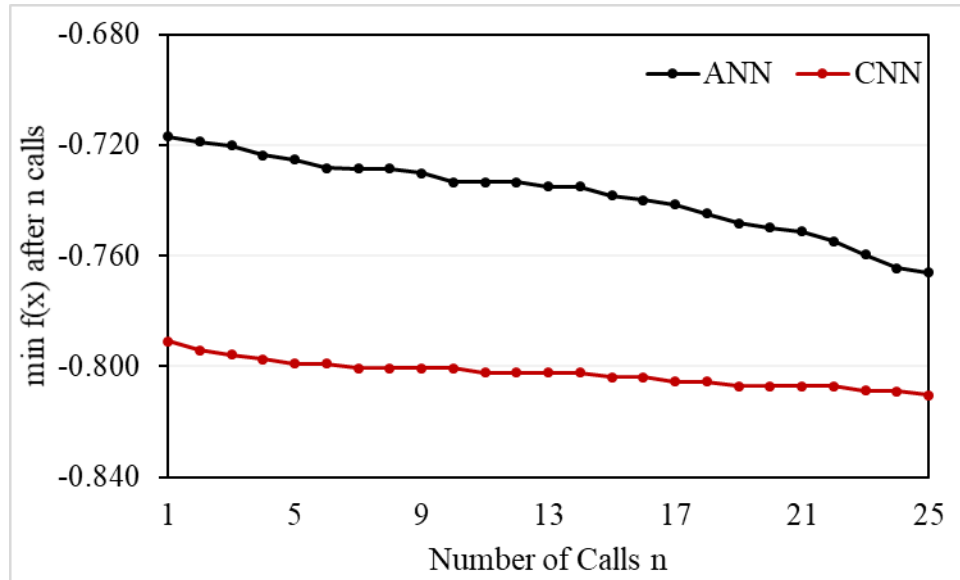


Fig. 8 The best-found configurations for the ANN and CNN models by Bayesian optimization. $f(x)$ represents the objective function (negative minimum AUROC with 5-fold cross validation).

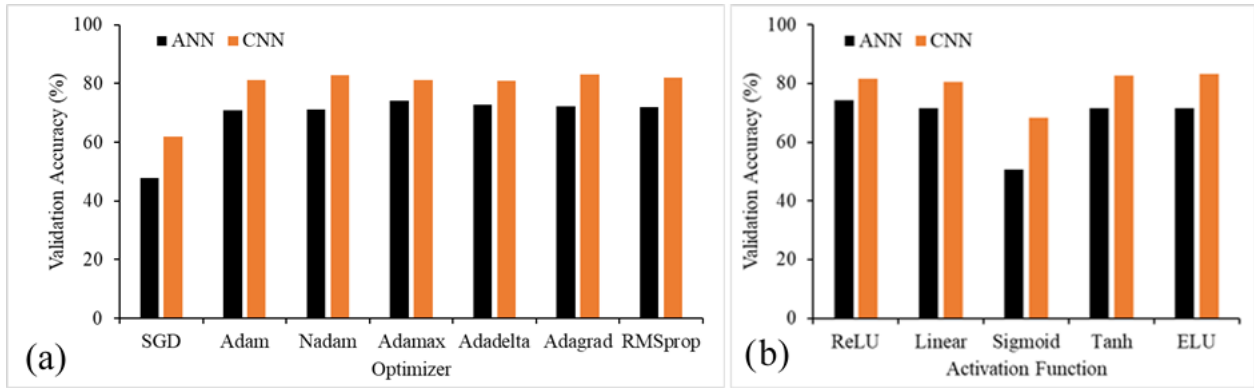


Fig.9 The effects of optimizer and activation function on the performance of the ANN and CNN models.

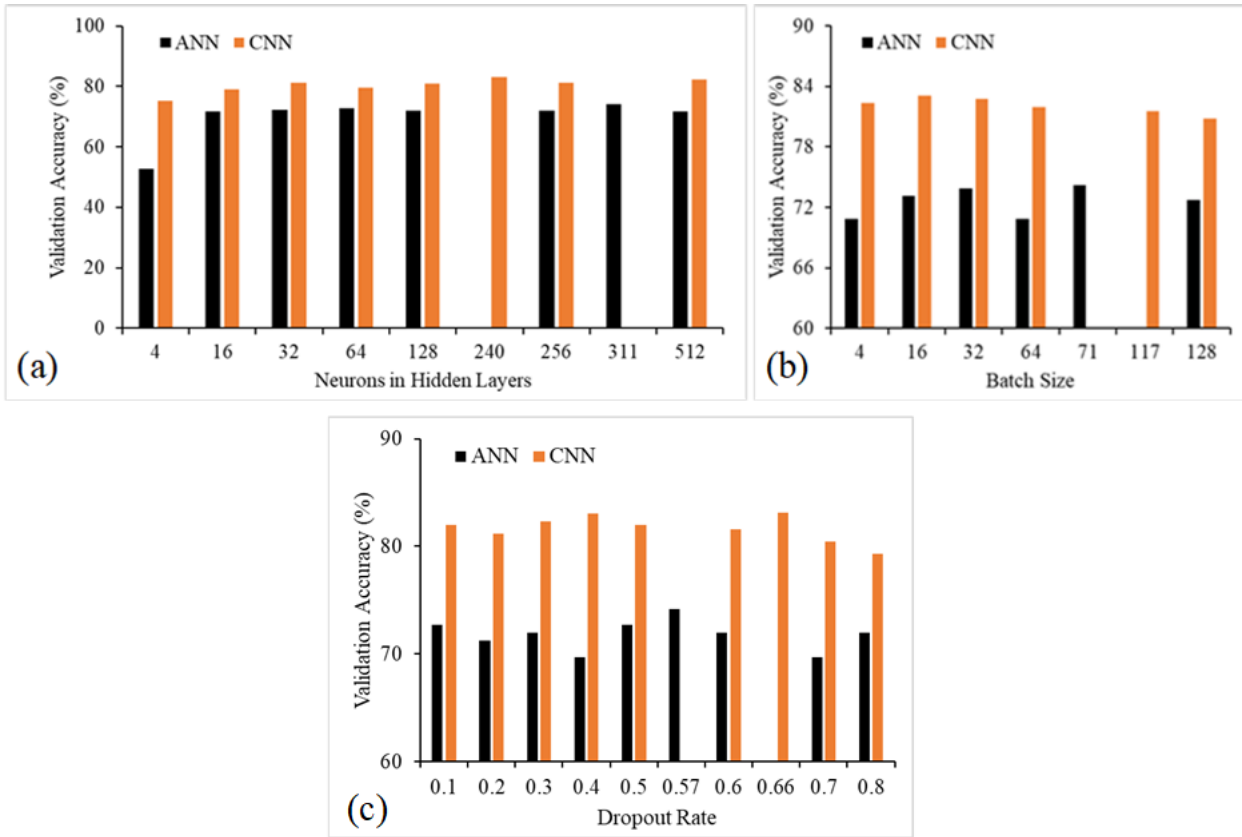


Fig. 10 The effects of the number of neurons in the hidden layers, batch size and dropout rate on the performance of the ANN and CNN models.

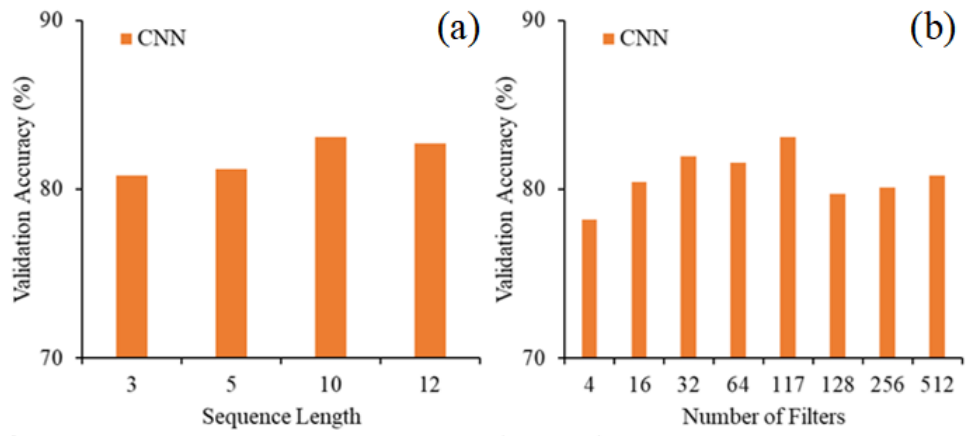


Fig. 11 The effects of the sequence length and the number of filters on the performance of the CNN model.

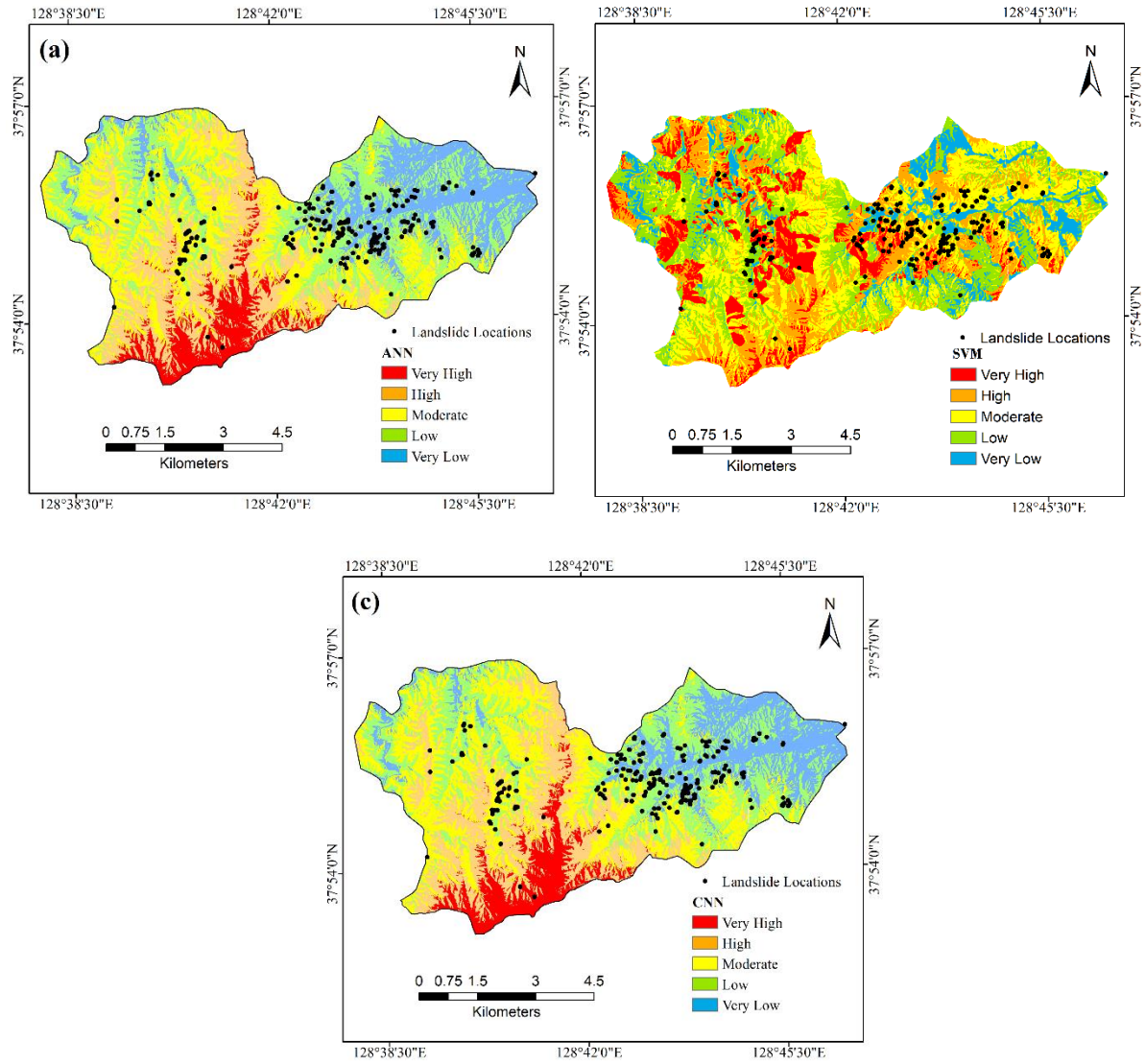


Fig. 12 Landslide susceptibility maps at the south of Yangyang using (a) ANN, (b) SVM, and (c) CNN.

Table 1 Landslide conditioning factors used in this research.

Original data	Factors	Data type	Scale
Aerial photograph ^a	Landslide location	Point	1:1,000
Topographical map ^b	Altitude	GRID	1:5,000
	Slope gradient		
	Slope aspect		
	Total curvature		
	Plan curvature		
	Profile curvature		
	Stream Power Index (SPI)		
	Sediment Transport Index (STI)		
	Topographic Roughness Index (TRI)		
	Topographic Wetness Index (TWI)		
Vegetation map ^c	Timber type	Polygon	1:25,000
	Timber density		
	Timber diameter		
	Timber age		
Geology map ^d	Lithology	Polygon	1:25,000
Soil map ^e	Soil texture	Polygon	1:25,000
Land use ^f	Land use type	Polygon	1:25,000

^a <http://map.daum.net>

^b National Geographic Information Institute (NGII)

^c Korea Forest Service (KFS)

^{d, e, f} Rural Development Administration (RDA)

Table 2 The common kernel functions used with SVM.

Kernel	Equation	Kernel Parameters
Radial basis function (RBF)	$K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2)$	γ
Linear	$K(x_i, x_j) = x_i^T x_j$	-
Polynomial	$K(x_i, x_j) = (-\gamma x_i^T x_j + 1)^d$	γ, d
Logistic	$K(x_i, x_j) = \text{Tanh}(-\gamma x_i^T x_j + 1)^d$	γ

Table 3 Hyperparameters and their search space of the proposed models.

Model	Parameters	Search Space
ANN	Batch Size	(4, 128)
	Activation Function	[<i>ReLU, Linear, Sigmoid, Tanh, ELU</i>]
	Optimization Method	[<i>SGD, Adam, Nadam, Adamax, Adadelta, Adagrad, RMSprop</i>]
	Neurons in Hidden Layers	(4, 512)
	Dropout Rate	(0, 0.8)
SVM	Kernel Function	[<i>RBF, Linear, Logistic, Polynomial</i>]
	C Value	(1, 500)
	Gamma	<i>Log space (-9, 3, 13)</i>
CNN	Number of Filters	(4, 512)
	Sequence Length	[3, 5, 10, 12]
	Batch Size	(4, 128)
	Activation Function	[<i>ReLU, Linear, Sigmoid, Tanh, ELU</i>]
	Optimization Method	[<i>SGD, Adam, Nadam, Adamax, Adadelta, Adagrad, RMSprop</i>]
	Neurons in Hidden Layers	(4, 512)
	Dropout Rate	(0, 0.8)

Table 4 Random Forest importance values and ranks estimated for the landslide conditioning factors. TRI (terrain roughness index), LULC(land use and land cover), STI (sediment transport index), SPI (stream power index), and TWI (terrain wetness index).

Landslide causal Factor	Rank	Importance
*Altitude	100	1.000
*Soil	87	0.869
*Plan Curvature	82	0.815
*TRI	80	0.802
*Total Curvature	76	0.759
*Lithology	76	0.757
*LULC	69	0.692
*STI	65	0.652
*SPI	65	0.650
*Profile Curvature	65	0.648
*TWI	65	0.647
*Slope	64	0.639
Timber Type	53	0.534
Timber Age	53	0.532
Aspect	53	0.530
Timber Diameter	35	0.353
Timber Density	27	0.269

* factors are significant at $\alpha = 0.05$.

Table 5 The best-selected hyperparameter values of different models.

Model	Parameters	Search Space
ANN	Batch Size	71
	Activation Function	<i>ReLU</i>
	Optimization Method	<i>Adamax</i>
	Neurons in Hidden Layers	311
	Dropout Rate	0.57
SVM	Kernel Function	RBF
	C Value	397
	Gamma	0.193
CNN	Number of Filters	117
	Sequence Length	10
	Batch Size	16
	Activation Function	<i>ELU</i>
	Optimization Method	<i>Adagrad</i>
	Neurons in Hidden Layers	240
	Dropout Rate	0.66

Table 6 Performance of the proposed deep learning models and the benchmark methods using the training and validation datasets.

Model	Training Accuracy (%)	Testing Accuracy (%)	Training AUC	Testing AUC	5-CV AUC
ANN	82.60	74.00	0.903	0.797	0.792
SVM	76.40	76.50	0.858	0.808	0.807
CNN	83.50	83.11	0.881	0.880	0.893

Table 7 Computing performance of the models in seconds.

Model	Computing Time (seconds)		
	Optimization	Training	Prediction ($\times 10^{-3}$)
ANN	2378.89	0.22	1.12
SVM	9.27	0.08	1.09
CNN	3408.83	0.59	2.92