Article

Landslide Susceptibility Modeling Using a Deep Random Neural Network

Cheng Huang¹,²,³, Fang Li⁴, Lei Wei²,³,* , Xudong Hu⁵ and Yingdong Yang²,³

¹ School of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650093, China
² Key Laboratory of Geohazard Forecast and Geoeological Restoration in Plateau Mountainous Area, MNR, Kunming 650216, China
³ Yunnan Institute of Geological Environment Monitoring, Kunming 650216, China
⁴ Land and Resources Information Center, Department of Natural Resources of Yunnan Province, Kunming 650224, China
⁵ School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

* Correspondence: ynsdzhjcywl@163.com

Abstract: Developing landslide susceptibility modeling is essential for detecting landslide-prone areas. Recently, deep learning theories and methods have been investigated in landslide modeling. However, their generalization is hindered because of the limited size of landslide data. In the present study, a novel deep learning-based landslide susceptibility assessment method named deep random neural network (DRNN) is proposed. In DRNN, a random mechanism is constructed to drop network layers and nodes randomly during landslide modeling. We take the Lushui area (Southwest China) as the case and select 12 landslide conditioning factors to perform landslide modeling. The performance evaluation results show that our method achieves desirable generalization performance (Kappa = 0.829) and outperforms other network models such as the convolution neural network (Kappa = 0.767), deep feedforward neural network (Kappa = 0.731), and Adaboost-based artificial neural network (Kappa = 0.732). Moreover, the robustness test shows the advantage of our DRNN, which is insensitive to variations in training data size. Our method yields an accuracy higher than 85% when the training data size stands at only 10%. The results demonstrate the effectiveness of the proposed landslide modeling method in enhancing generalization. The proposed DRNN produces accurate results in terms of delineating landslide-prone areas and shows promising applications.

Keywords: landslide; deep learning; neural network; machine learning

1. Introduction

A landslide is a geo-hazard that results in slope displacement. Many landslide events occur annually because of negative factors, e.g., weak rocks, rainfall, and human activities [1]. Over recent years, landslides have caused severe damage to the socio-economic environment due to the loss of lives and property [2]. Thus, exploring the landslide prediction method is necessary to prevent the harm of landslides. Landslide susceptibility mapping (LSM) is a useful method for evaluating which places are subject to landslides. The landslide susceptibility index (LSI) from the LSM can indicate the probability of landslide occurrence spatially. Therefore, landslide susceptibility modeling is usually treated as the prerequisite of landslide prevention and management.

Traditional evaluation methods mainly rely on knowledge and experience to determine landslide-prone areas. These methods are time-consuming and labor-intensive, and their results are easily affected by individual subjectivity [3]. Data-driven methods, particularly machine learning methods, have become increasingly popular in recent years. Data-driven methods directly estimate landslide spatial occurrence probability according to landslide-related factors. Therefore, data-driven landslide modeling methods do not need to consider
the detailed mechanical process of slope sliding but rely on model structure and input data [4]. However, the relationship between landslide-triggering factors and landslides is complex, thereby complicating landslide susceptibility modeling.

Numerous methods have been developed for landslide susceptibility modeling by exploring the relationship between landslides and their related factors. Early research mainly adopted statistics-based data models, including the grey model [5], exponential smoothing model [6], statistical index [7], frequency ratio [8], analytical hierarchy process [9], and weight of evidence [10]. These methods effectively deal with landslide prediction under simple situations or specific conditions. However, they have insufficient accuracies in complex and variable patterns. Given the rapid development of computer techniques, machine learning methods have been widely investigated for LSM. Commonly used machine learning methods include the maximum entropy [11], artificial neural network (ANN) [12], Bayesian network [13], support vector machine (SVM) [14], kernel logistic regression [15], boosted regression tree [16], and logistic model tree [17]. These algorithms treat landslide-influencing factors as independent variables to build the non-linear relationship between factors and landslides. Besides the landslide spatial prediction, the success of machine learning is exhibited in other geotechnical engineering problems, such as the prediction of the soil compression index [18], flood vulnerability analysis [19], and earthquake early warning [20]. However, single machine learning methods tend to suffer from local optimization problems due to the limited hypothesis space [21]. Therefore, scholars have explored ensemble learning methods for LSM [22]. Many ensemble algorithms have been developed and proposed for LSM, such as Bagging [23], Boosting [24], random forest (RF) [25], rotation forest [26], stacking [27], ANN-Bayes analysis [28], GWO-SVM [29], and random subspace-based naive Bayes tree [30]. Ensemble methods can expand the hypothesis space of models to search for the optimal solution; thus, their performance is better than that of single models [31,32].

Recently, deep learning methods have achieved state-of-the-art results in LSM studies [33–35]. The existing deep learning-based LSM techniques are mainly inspired by the deep feedforward neural network (DNN) [36], recurrent neural network [37], convolution neural network (CNN) [38], and auto-encoder [39] (Table 1). Deep learning ensemble methods have been developed in order to improve the prediction capability of individual deep learning models [40–45]. Several neuron-free deep learning methods are also successfully applied in LSM, such as the deep Boosting model and the deep learning tree [44]. Compared with shallow machine learning methods, deep learning methods can further improve landslide prediction accuracy due to the deep and interconnected structures [45]. In particular, the CNN plays an important role in addressing feature learning problems [33]. The CNN is a neural network that simulates the mechanism of the human visual system. The core feature encoding the structure of the CNN is the convolution filter, which establishes local connections for the input data. By using stacking convolution and pooling layers, CNN can extract robust and diverse representations [46]. The convolution-based deep learning models have been widely applied to various fields, including image classification and segmentation [38], video summarization [47], remote sensing [48], object detection [49], and natural language processing [50]. CNN also brings a revolution to landslide susceptibility modeling, which can mine deep-level relationships between landslide-related factors and landslides and explore the synergic effects of factors on landslides through convolution operators [33]. In LSM, CNNs are found to be superior to shallow machine learning methods. For example, Liu et al. (2022) compared the CNN with three conventional machine learning models of RF, logistic regression, and SVM. They showed that CNN not only achieves preferable performance but also reduces the salt-and-pepper effect [51].
Table 1. A summary of the existing deep learning-based LSM methods.

<table>
<thead>
<tr>
<th>Basic structures</th>
<th>Method Descriptions</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep feedforward neural network (DNN)</td>
<td>• A DNN model with greedy unsupervised learning</td>
<td>[52]</td>
</tr>
<tr>
<td></td>
<td>• SSL-DNN: A novel DNN model with semi-supervised learning</td>
<td>[53]</td>
</tr>
<tr>
<td></td>
<td>• A kernel-based DNN</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td>• ACO-DBN: The DNN was integrated with the ant colony optimization strategy</td>
<td>[55]</td>
</tr>
<tr>
<td></td>
<td>• A robust DNN model based on the combination of a DNN, extreme learning machine, ANN, and genetic algorithm</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td>• A DNN model with the particle swarm algorithm</td>
<td>[57]</td>
</tr>
<tr>
<td></td>
<td>• A DNN model integrated with the SVM</td>
<td>[58]</td>
</tr>
<tr>
<td>Convolution neural network (CNN)</td>
<td>• CNN-1D, CNN-2D, and CNN-3D methods</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td>• The CNN-1D with the Bayesian optimization algorithm</td>
<td>[46]</td>
</tr>
<tr>
<td></td>
<td>• The CNN integrated with SVM, RF, and logistic regression</td>
<td>[59]</td>
</tr>
<tr>
<td></td>
<td>• The CNN integrated with the metaheuristic optimization algorithm</td>
<td>[60]</td>
</tr>
<tr>
<td></td>
<td>• LSGH-Net: A novel CNN method based on the CNN and Gaussian heatmap sampling technique</td>
<td>[61]</td>
</tr>
<tr>
<td>Recurrent neural network (RNN)</td>
<td>• A typical RNN model</td>
<td>[62]</td>
</tr>
<tr>
<td></td>
<td>• A long-short-term memory (LSTM) model</td>
<td>[63]</td>
</tr>
<tr>
<td></td>
<td>• A novel RNN model combining LSTM with a conditional random field</td>
<td>[64]</td>
</tr>
<tr>
<td>Auto-encoder</td>
<td>• A typical auto-encoder model</td>
<td>[65]</td>
</tr>
<tr>
<td></td>
<td>• A fully connected sparse auto-encoder model</td>
<td>[66]</td>
</tr>
<tr>
<td></td>
<td>• An auto-encoder model integrated with the RF</td>
<td>[67]</td>
</tr>
<tr>
<td></td>
<td>• A stacked auto-encoder with the sparse optimization</td>
<td>[68]</td>
</tr>
<tr>
<td>Deep learning ensemble methods</td>
<td>• An ensemble model of CNN and RNN based on a stacking algorithm</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>• An ensemble model of CNN, RNN, and LSTM based on shared blocks</td>
<td>[41]</td>
</tr>
<tr>
<td></td>
<td>• An ensemble model of CNN, DNN, and ResNet</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>• An ensemble model of CNN and RNN variants</td>
<td>[43]</td>
</tr>
<tr>
<td>Other methods</td>
<td>• A deep boosting method</td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td>• A deep learning tree</td>
<td>[44]</td>
</tr>
</tbody>
</table>

Although deep learning methods achieve impressive performance, they rely on massive sample data for training and learning [69]. Given that the LSM training data are relatively limited, the designed model should not be complex, e.g., decreasing the depth of
the network. The reason is that the size of the limited data does not match the volumes of the parameters of the complex structure, whereas the complex structure may lead to serious overfitting problems when the training data are insufficient. Therefore, the performance of deep learning methods is largely constrained in landslide susceptibility modeling due to the limitation of training data. In machine learning-based landslide susceptibility modeling, ensemble methods that rely on random strategies are usually effective in dealing with limited data. For example, Bagging randomly extracts subsets of samples to estimate landslide susceptibility [22]; rotation forest [70], random subspace [30], and RF adopt random features to improve the model’s generalization skill. In wetland classification, Hu et al. (2021) showed that the ANN with rotation forest effectively overcomes the negative impacts of reducing the training data size [32]. Similar techniques, such as DropOut [71], StochasticDepth [72], DropPath [73], and DropBlock [74] are also developed in deep learning. The basic idea of these methods is to insert perturbation into the model’s structure randomly to avoid overfitting the training data. For instance, in DropPath, the network’s sub-paths in the residual blocks of the FractalNet are randomly removed [73]. In landslide susceptibility modeling, deep learning methods require perturbation to be regularized. Therefore, exploring deep learning landslide models with random strategies can potentially handle limited landslide data.

The present study proposes a novel deep learning-based method, named deep random neural network (DRNN), to improve the generalization capability of landslide modeling. DRNN is built based on convolution layers and fully connected layers to learn deep features. DRNN is characterized by a random mechanism that randomly drops a subset of the network’s layers and nodes during the training phase. This design helps the model reduce the reliance on training data and allows the model to achieve a deep structure while avoiding overfitting. To evaluate the effectiveness of the DRNN, the Lushui county in Southwest China is taken as the case area. Furthermore, three networks, namely, CNN, DNN, and Adaboost-based ANN (ADAANN), were implemented for method comparisons. Finally, we adopt the proposed DRNN to generate the landslide susceptibility map of the study area.

2. Methodology

2.1. Whole Structure of the DRNN

The whole DRNN structure consists of a random input layer, a 1D convolution layer, a max-pooling layer, a random hidden block, and an output layer (Figure 1). The convolution layer and the random hidden block are fundamental components of the DRNN. The convolution layer is used to explore the deep-level and synergic effects of landslide-related factors on landslides. Since the small size of the landslide data makes fully training the model with deep architectures difficult, we introduce a random strategy into the DRNN. The random strategy is applied to the nodes of network layers and the layer depth. When the 1D sequence factors are input to convolution layers, a part of the factors is randomly taken to form a factor subspace for signal passing. This mechanism is similar to the processes of RF and random subspace, which allow the model to receive diverse information. Furthermore, when signals are propagated to fully connected layers via convolution layers, a certain number of hidden layers is randomly dropped. In this case, the connections of the dropped layers with their neighbor layers are canceled. Meanwhile, the nodes in the remaining hidden layers are processed using the same scheme as the nodes in the input layer.
The present study proposes a novel deep learning-based method, named deep random neural network (DRNN), to improve the generalization capability of landslide modeling. DRNN is built based on convolution layers and fully connected layers to learn deep features. DRNN is characterized by a random mechanism that randomly drops a subset of the model’s layers and nodes during the training phase. This design helps the model reduce the reliance on training data and allows the model to achieve a deep structure.

2.2. Random Input Layer

The random input layer of the DRNN includes several nodes, representing landslide conditioning factors. Let \( X = \{ x_1, x_2, \ldots, x_N \} \) be the input 1D sequence factor, where \( N \) denotes the number of factors. In each iteration, a random binary mask \( M_{\text{input}} \in \mathbb{R}^{N \times 1} \) is generated, where the value of a certain number of locations in the \( M_{\text{input}} \) is set to zero and the remaining is one. The nodes in the input layer are first processed by the \( M_{\text{input}} \):

\[
X' = X \times M_{\text{input}} \quad (1)
\]

The masking ratio of \( M_{\text{input}} \) is controlled by a defined parameter \( R_{\text{node}} \in [0,1] \). The nodes being masked means that the corresponding landslide conditioning factor is not considered for modeling in the current iteration. Note that the node masking is only utilized in training.

2.3. Convolutional and Max-Pooling Layers

Like CNN, DRNN uses the convolution layer and max-pooling layer to learn local relationships between landslide feature vectors. By using the convolution operation, each element in the sequence is connected to its \( K \) neighborhoods. The size of the convolution filter is assumed to be \( K \times 1 \times D \). In this case, the signal is processed by convolution operation and is then nonlinearly activated. The process of convolution is expressed as:

\[
C_i = \sum_{n}^{N} f(\omega_i \ast x_n + b_i) \quad i \in [1,D], \; n \in [1,N] \quad (2)
\]

where, \( C_i \) and \( x_n \) represent the output and input of the convolution layer, respectively. \( \omega_i \) and \( b_i \) are weights and biases, respectively. \( f(\cdot) \) is the activation function. In this study, the ReLU function is used for the nonlinear transformation of signals.

\[
\text{ReLU}(x) = \begin{cases} 
  x & x \geq 0 \\
  0 & x < 0 
\end{cases} \quad (3)
\]

The aim of the pooling operation is to reduce parameters and increase the receptive field of the model. Assuming the max-pooling filter has a size of \( p \times 1 \), the max-pooling layer integrates the convoluted features into a new sequence feature with a size of \( \frac{N-K+1}{p} \times 1 \).

\[
P_i = \max_p(C_i) \quad (4)
\]

where \( P_i \) is the output of the max-pooling layer, which contains the main important information of \( C_i \). \( \max_p(\cdot) \) is the max-pooling function with a pooling size of \( p \times 1 \).
2.4. Random Hidden Block

The landslide sequence factors extracted from the max-pooling layer and the convolution layer are sent to the random hidden block, which is a modified full connection layer. Since the DRNN contains deep structures that may cause overfitting and gradient disappearance problems, we propose a hidden-layer-dropping strategy to encourage regularization. Specifically, we randomly drop several hidden layers and bypass their locations using skip connections during the model training stage. In this case, the dropped layers are inactivated, and their connections with adjacent layers are interrupted. The two adjacent layers of the dropped layer are still available because the skip connection technique establishes a new path to link them, guaranteeing the normal operation of signal propagations. Assuming that the random hidden block has \( L \) hidden layers, the function \( F_l(\cdot) \) represents the original mapping between the \((l - 1)\)-th and \( l \)-th layers, where \( l \in [1, L] \). Let \( v_l \in \{0, 1\} \) be a binary variable which indicates whether the \( l \)-th hidden layer is activated \((v_l = 1)\) or inactivated \((v_l = 0)\).

\[
H_l = v_l \times F_l(H_{l-1}) + (1 - v_l) \times I(H_{l-1})
\]

where \( H_{l-1} \) and \( H_l \) are the input and output of the \( l \)-th hidden layer. In this study, the number of the dropped layers is controlled by a defined parameter \( R_{layer} \in [0, 1] \). \( I(\cdot) \) denotes identity mapping, indicating that the output of a layer maintains the same values as the input. The original nonlinear mapping from layer \( l - 1 \) to \( l \) is reduced to identity mapping when the \( l \)-th hidden layer is dropped. This propagation rule is combined with the skip connection strategy, allowing the network to propagate signals back and forth between adjacent or nonadjacent layers.

During the model inference stage, all layers in the network are activated in order to employ the full capability of the model. The gradient propagation process during inference is expressed as follows:

\[
H_l = F_l(H_{l-1})
\]

As for the nodes in the activated hidden layers, they are randomly dropped using a ratio \( R_{node} \) throughout training. Similarly, we create a random binary mask \( M_{hidden} \in R^{N - K + 1 \times 1} \) to achieve this process:

\[
H_l' = H_l \times M_{hidden}
\]

The random strategy of the DRNN can be viewed as injecting noise into the network structure (e.g., nodes of network layers and layer depth), thereby allowing the model to learn robust and generalized representations from limited training data. Due to a random mechanism, the structures involved in the training at each iteration are different. From the ensemble learning perspective, training a single DRNN model with random nodes and random layers can be interpreted as constructing an implicit model ensemble.

2.5. Output Layer

The final output layer has two nodes representing the landslide (1) and non-landslide (0). The prediction of the DRNN is made by the SoftMax classifier [38]. SoftMax is a normalized exponential function, which is expressed as follows:

\[
\text{SoftMax}(x) = \frac{x_k}{\sum x_k}
\]

where \( x_k \) represents the \( k \)-th element of the output and \( k \) belongs to \{0, 1\}.

3. Study Area and Data Used

3.1. Study Area

The Lushui area belongs to Yunnan Province in Southwest China. It lies between 98°34′ and 99°09′ east longitude and 25°33′ to 26°32′ north latitude (Figure 2). The area has an elevation ranging from 740 m to 4160 m. The whole region is crossed by the Nujiang River Grand Canyon. Due to the special terrain, nearly half of the area has a slope higher
than 30°. The rainy season generally starts in May and ends in October. The rainfall during this stage occupies approximately 70% of the total annual rainfall.

Figure 2. Position of the study area.

The geology units of Lushui county vary from Precambrian to Quaternary. The Lushui area is widely occupied by metamorphic, igneous, and sedimentary rocks and unconsolidated sediments (Figure 3). The oldest geology units majorly consist of the Proterozoic Gaoligongshan (Ptgl) and Chongshan (Ptc) groups, and the youngest geological unit is Quaternary. According to previous studies [27], the geological unit of the area is classified into seven engineering rock groups (ERGs).
Under the combined influence of geological, topographical, and hydrologic factors, the stress acting on the slope easily exceeds the limitation. In the study area, the slope-forming materials mainly consist of alluvial and pluvial deposits, which are widely distributed along the Nujiang River. Complex environments make the Lushui area prone to landslides. It is an urgent task to perform accurate and reliable LSM for the study area.

3.2. Landslide Inventory Map

Landslide inventory records important information about landslides such as location, size, and type. Therefore, preparing a landslide inventory is essential. According to the extensive field surveys made by the Department of Nature Resources of Yunnan Province, 413 landslides are found in the Lushui area (Figure 4a). The data collection time is from 2014 to 2020. The dominant type of landslides is soil landslides, accounting for 90% of the total landslides. Among these landslides, the smallest, largest, and average areas are 80 m$^2$, 600,000 m$^2$, and 39,557 m$^2$, respectively. According to the landslide volume, 53% of landslides are classified as small landslides, followed by medium landslides (41.4%), large landslides (5.3%), and huge landslides (0.3%). These landslides are easily affected by rainfall and excavation activities.

To conduct landslide modeling, it is suggested to adopt 70% of the total landslides (289 landslides) for training, and the remaining 30% landslides (124 landslides) are used for testing. To select reliable and representative non-landslide locations from landslide-free areas, the K-means clustering method was adopted [27]. The whole study area was converted to grid units with 30 m × 30 m, and the non-landslide units were classified into 413 abstract categories through K-means clustering. The units closest to the center of each category were chosen as negative samples (Figure 4b).

3.3. Landslide Conditioning Factors

The selection of landslide conditioning factors is important in LSM. This study mainly consults the work of Hu et al. (2020), who performed a detailed analysis of landslide-
influencing factors in the study area [27]. The only difference is that the plan curvature was replaced with the topographic wetness index (TWI) because the plan curvature makes little contribution to landslide prediction. Accordingly, a total of 12 factors which are associated with landslide occurrence were considered for modeling landslide susceptibility in this study. These factors include elevation, slope angle, slope aspect, profile curvature, TWI, ERG, land use, distance to roads, distance to rivers, distance to faults, rainfall, and normalized difference vegetation index (NDVI). The selected factors comprehensively reflect the topographical, geological, hydrological, and anthropogenic features of the study area.

The digital elevation model (DEM) data were first prepared using a digital contour map at a scale of 1:10000. The DEM was converted into the raster format with a resolution of 30 m. The terrain factors, including the elevation, slope angle, slope aspect, and profile curvature, were extracted from the DEM. TWI is also a commonly used factor that indicates the influence of terrain on soil runoff processes, which is calculated as follows:

\[ \text{TWI} = \ln \left( \frac{\alpha}{\tan \beta} \right) \]  

where \( \alpha \) is the cumulative catchment area, and \( \beta \) is the slope angle.

The distance to roads, distance to rivers, and land use were collected according to the Third Detailed Land Investigation Nationwide (China) at the county scale. The distance to roads and the distance to rivers range from 0 m to 9993 m and 0 m to 2567 m, respectively. The land use map was reclassified into seven main types: residential area, forest, grassland, farmland, bare land, engineering land, and other lands.

Geological factors can never be neglected in analyzing landslide occurrence probability [75]. The ERG and distance to faults are extracted from a 1:100,000-scale geological map. The ERG of the study area includes seven groups: the group of layered, massive hard metamorphic rock; the group of layered, flaky soft mudstone, shale, and sandstone; the group of layered karst medium hard carbonate rock; the group of layered hard sandy mudstone; the group of massive, vein hard extrusive rock; the group of massive hard intrusive rock; the group of loose and semi-cemented rock dominated by gravel and sand. The range of the distance to faults is 0–12,192 m.

The NDVI is an important indicator that reflects the relationship between landslides and vegetation [51]. The NDVI was calculated based on the 30 m × 30 m Landsat 8 OLI remote sensing image from the USGS website (https://earthquake.usgs.gov/). The remote sensing image was acquired on 7 March 2018. The calculation of the NDVI is related to the red band and infrared band:

\[ \text{NDVI} = \frac{\text{Band}_{IR} - \text{Band}_{R}}{\text{Band}_{IR} + \text{Band}_{R}} \]

where \( \text{Band}_{IR} \) and \( \text{Band}_{R} \) represent the red band and infrared band, respectively.

Rainfall is a critical factor contributing to slope movement. We collected the annual average rainfall data of all villages from the Yunnan Digital Village Website (http://www.ynszxc.net). The data was collected on 19 September 2018. Then, we produced an annual rainfall map using the Kriging interpolation method.

All vector layers were transformed into 30 m × 30 m raster data in order to unify the data format (Figure 5). To make the data structure suitable for the DRNN, the stacking factors were reshaped to 1D sequence factors in a tensor format when input to the DRNN.
Figure 5. Cont.
Figure 5. Landslide conditioning factors: (a) Elevation, (b) Slope angle, (c) Slope aspect, (d) Profile curvature, (e) TWI, (f) ERG, (g) Land use, (h) Distance to roads, (i) Distance to rivers, (j) Distance to faults, (k) Annual rainfall, (l) NDVI.

4. Results and Discussion

To train the DRNN, the adaptive moment estimation [76] optimizer and the cross-entropy loss were adopted. The learning rate was set to 0.001, which was decreased by a factor of 0.7 every 1/4 epochs. The DRNN was trained with 200 epochs. The sizes of the convolution filter and the max-pooling filter were set to $3 \times 1 \times 64$ and $2 \times 1$. The default numbers of hidden layers $l$, the node dropping ratio $R_{node}$, and the layer dropping ratio $R_{layer}$ were set to 16, 0.5, and 0.5, respectively. The goal of training DRNN is to optimize the discrepancy between the predicted and real values. Figure 6 presents the variation in the loss during the training stage. With the increase in the epoch, the training loss and validating loss reduce until both loss values converge at a constant level, suggesting a satisfactory training process.

Figure 6. The variation of loss values during the training phase.

Furthermore, two deep learning networks, namely, CNN and DNN, and an ensemble-based network, ADAANN, were implemented for method comparisons. CNN is a biologically inspired neural network originally proposed for image classification [77]. CNN mainly benefits from local connections, shared weights, and pooling operations for feature
learning. DNN is a deep feedforward network with interconnected neurons, which can be regarded as the ANN with multiple hidden layers. ADAANN is a hybrid model of Adaboost and ANN, where the Adaboost algorithm enhances the fitting power of the ANN through cost-sensitive ideas.

### 4.1. Landslide Susceptibility Mapping

On the basis of landslide conditioning factors, landslide models estimate the LSI for each unit of the study area. The LSI values were reclassified into five susceptibility levels, namely, very low, low, moderate, high, and very high, by using the geometrical interval classification method. The resulting landslide susceptibility maps using various models are shown in Figure 7. These models yield a similar landslide susceptibility distribution pattern, that is, highly susceptible areas are usually distributed along the Nujiang River Grand Canyon. The results are highly consistent with the true state of the study area because these landslide-prone areas are easily affected by tectonic movements, water erosion, and road construction. Loose soil and soft rock are further conducive to slope failure under low vegetation coverage and frequent rainfall.

The detailed statistic results on landslide susceptibility distribution are exhibited in Table 2. The DRNN model predicts that 16.66% of the whole area shows very high landslide susceptibility, whereas 19.06%, 11.08%, 19.05%, and 34.15% of the area have high, moderate, low, and very low susceptibility levels, respectively. CNN predicts that 16.98% of the area has a landslide occurrence probability, whereas 23.46%, 10.36%, 9.04%, and 40.17% of the total area lie in high, moderate, low, and very low susceptibilities, respectively. Regarding the DNN, 66.59% of the landslides are located in a very high class that accounts for 18.22% of the whole study area. As for DRNN, 68.04% of the landslides fall into a very high vulnerability class, while these regions account for 18.57% of the whole region.

Additionally, the reliability of LSM was quantitatively measured using the landslide density (LD) index, which is a proportion of PL (percentage of landslide pixels) and PC (percentage of pixels in a susceptibility class). Table 2 shows that the LD of various landslide models consistently increases as the susceptibility level improves. The results indicate that these network models achieve reasonable landslide susceptibility assessments. Particularly, the proposed DRNN achieves the highest LD value, suggesting that it may be highly appropriate for predicting landslide susceptibility.

### 4.2. Performance Evaluation and Comparison

In this study, the performance of the DRNN was evaluated using the overall accuracy (OA), kappa coefficient (K), F-measure (F), Matthews correlation coefficient (MCC), and receiver operating characteristic (ROC) curve. The performance evaluation was conducted using both the training data and the testing data (Table 3). On the basis of training data, the fitting power of models was measured. All models achieve a high level of goodness of fit. Note that the CNN shows higher values of OA, K, F, and MCC than DRNN, DNN, and ADAANN, proving the excellent feature extraction capability of the convolution structure. The proposed DRNN achieves the second-best fitting performance, yielding OA, K, F, and MCC values of 91.16%, 0.823, 0.912, and 0.823, respectively. By comparison, DNN and ADAANN obtain nearly equal performances in terms of goodness of fit. In terms of
generalization performance, our method achieves the highest OA, K, F, and MCC, with values of 91.46%, 0.829, 0.914, and 0.828, respectively. In addition, CNN also achieves satisfactory performance, with OA, K, F, and MCC values of 88.41%, 0.7767, 0.883, and 0.767, respectively. DNN and ADAANN perform comparably regarding the generalization ability. The ROC curves of various models are shown in Figure 8. The area under the ROC curve (AUC) quantitatively depicts the model’s performance. On the training data, CNN achieves the highest AUC (0.981). It is followed by the DRNN (0.967), DNN (0.940), and ADAANN (0.936). On the testing data, the proposed DRNN holds the best prediction performance in AUC (0.946), followed by CNN (0.942), DNN (0.936), and ADAANN (0.916).
Table 3. Performance evaluation and comparison of landslide models.

<table>
<thead>
<tr>
<th>Models</th>
<th>On the Training Data</th>
<th>On the Testing Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OA (%)</td>
<td>K</td>
</tr>
<tr>
<td>ADAANN</td>
<td>88.22</td>
<td>0.764</td>
</tr>
<tr>
<td>DNN</td>
<td>88.26</td>
<td>0.765</td>
</tr>
<tr>
<td>CNN</td>
<td>92.53</td>
<td>0.850</td>
</tr>
<tr>
<td>DRNN</td>
<td>91.16</td>
<td>0.823</td>
</tr>
</tbody>
</table>

Figure 8. ROC curves of different neural network models (a) on the training dataset, (b) on the validation dataset.

Similar results can be found in previous landslide studies. For example, Wang et al. (2022) assessed the landslide susceptibility of the Yaan-Linzhi area on the basis of deep learning methods [78]. They found that the accuracy of CNN is greater than that of the DNN in both the training and testing stages. Aslam et al. (2022) analyzed the performance of different CNN structures in landslide modeling and proved the advantages of CNN relative to DNN [79]. Nguyen and Kim (2021) compared deep learning and ensemble learning methods for landslide spatial probability prediction and concluded that CNN significantly outperforms DNN and ensemble models including RF and Adaboost [80]. In the present study, CNN achieves the best prediction accuracy on training data but encounters a slight overfitting problem. The proposed DRNN effectively alleviates this limitation by using the random mechanism. The random mechanism expands the data space, allowing the model to receive diverse information for full training. Additionally, the random rule acts as regularization, which helps mitigate overfitting. The DRNN achieves desirable and comprehensive performances, which may be suitable for landslide susceptibility assessment in the study area.

4.3. Robustness to Training Data Size

To fully understand the advantages of the proposed landslide susceptibility modeling method, the robustness to the training data size is examined (Figure 9). The robustness test is important because the number of landslides may be too small to train the deep learning models fully in some cases. The robustness test was implemented by changing the partition of training and testing data from 80%/20% to 10%/90%. This means that the amount of training data is gradually decreased while the amount of testing data is gradually increased during experiments. We took the OA and K as examples to visualize the variation in the model’s performance. It can be observed that the K values of the four network models show a downward tendency with the reduction in training data (from 80% to 10%), with
decimals 10.49%, 30.97%, 13.78%, and 11.02% for DRNN, DNN, CNN, and ADAANN, respectively. The OA values decline by 5.28%, 15.58%, 6.97%, and 5.51% for DRNN, DNN, CNN, and ADAANN, respectively. The DRNN is consistently better than other network models. Note that DRNN still achieves OA values higher than 85% when the remaining training data size is only 10%. The DRNN achieves the lowest standard deviation in terms of K, with a value of 3.06, followed by ADAANN (3.80), CNN (4.91), and DNN (9.23), respectively. Similarly, DRNN achieves the best standard deviation of OA, with a value of 1.53, followed by ADAANN (1.90), CNN (2.48), and DNN (4.63), respectively. These results demonstrate that the DRNN can keep robustness to the training data size effectively. The random strategy of the DRNN helps alleviate the negative influence of limited data on landslide modeling accuracy.

Figure 9. Robustness test on the training data size: (a) Changes in K values, (b) Changes in OA values.

4.4. Influence of Critical Parameter Settings

In this subsection, the impact of the important parameters of the DRNN on landslide modeling accuracy is studied. These parameters mainly include $L$ (the number of hidden layers in the random hidden block), $R_{\text{node}}$ (the node dropping ratio), and $R_{\text{layer}}$ (the layer dropping ratio).

In terms of hidden layers, all performance metrics increase first and then decrease as $L$ increases from 8 to 20, reaching the maximum value when $L = 16$ (Figure 10a). Increasing hidden layers means increasing the model’s complexity, which is conducive to the fitting power [81]. However, overemphasizing the fitting power may damage the generalization of landslide modeling. Thus, the number of hidden layers should be set to a moderate value. The random strategy of DRNN makes expanding the model’s depth to 20 layers feasible while avoiding overfitting. This development may be helpful for future landslide modeling studies exploring deeper network models.

Figure 10. Influence of critical parameters: (a) Hidden layers, (b) Node dropping ratio, (c) Layer dropping ratio.

The results with respect to $R_{\text{node}}$ and $R_{\text{layer}}$ are shown in Figure 10b,c, which present some similarities. The model variants of DRNN achieve the highest landslide modeling accuracies at moderate levels of $R_{\text{node}}$ and $R_{\text{layer}}$. The results are reasonable because
dropping too many nodes or layers may damage the model’s structure, whereas too slight of a perturbation is insufficient for realizing effective regularization.

5. Conclusions

Landslides are one of the most dangerous geology disasters; they usually cause serious damage to society and the environment. Thus, developing accurate landslide susceptibility modeling is necessary in order to perform early warning work. This study proposes the DRNN, a novel deep learning-based landslide model, for landslide susceptibility assessment. The DRNN randomly drops network layers and nodes to learn generalized features during landslide modeling. With the Lushui area as the case, the performance evaluation shows that the DRNN achieves the highest generalization accuracy (OA = 91.46%) and outperforms other network models, such as CNN (OA = 88.41%), DNN (OA = 86.58%), and ADAANN (OA = 86.59%). DRNN also achieves robust performance due to the random mechanism, which is insensitive to variations in training data size. Thus, our method overcomes the limitation of the layer depth of deep learning in landslide modeling, thereby effectively mitigating overfitting and enhancing the generalization capability. In our case study, the proposed DRNN produces the most accurate and suitable results for detecting landslide-vulnerable areas, which shows promising applications. In the future, we will explore the effectiveness of our method in other scenes.

Author Contributions: C.H. and L.W. suggested the idea of modeling landslide susceptibility and organized this paper. C.H. and X.H. implemented deep learning models. C.H., F.L. and Y.Y. collected the landslide data and processed the landslide-influencing factors. All of the authors contributed to manuscript editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by the Construction of Meteorological Risk Early-Warning Service for Geological Disasters in Yunnan Province (No. 2022004); the Information Construction of Geological Environment in Yunnan Province (No. 2019055); and the Application Study of Key Technologies of Spatial-temporal Information Service for Rain-induced Landslides (No. 2022068).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

3. Xing, Y.; Yue, J.P.; Chen, C. Interval Estimation of Landslide Displacement Prediction Based on Time Series Decomposition and Long Short-Term Memory Network. IEEE Access 2020, 8, 3187–3196. [CrossRef]


12. Zare, M.; Pourghasemi, H.R.; Vafakhah, M.; Pradhan, B. Landslide Susceptibility Mapping at Vaz Watershed (Iran) Using an Artificial Neural Network Model: A Comparison between Multilayer Perceptron (MLP) and Radial Basic Function (RBF) Algorithms. *Arab. J. Geosci.* 2013, 6, 2873−2888. [CrossRef]


