remote sensing

MDPI

Article

A Combination of Deep Autoencoder and Multi-Scale Residual Network for Landslide Susceptibility Evaluation

Zhuolu Wang [1], Shenghua Xu [2,*], Jiping Liu [2], Yong Wang [2], Xinrui Ma [1], Tao Jiang [3], Xuan He [1] and Zeya Han [1]

1 School of Mapping and Geographical Science, Liaoning Technical University, Fuxin 123000, China
2 Chinese Academy of Surveying and Mapping, Beijing 100830, China
3 School of Resources and Environmental Sciences, Wuhan University, Wuhan 430072, China
* Correspondence: xushh@casm.ac.cn

Abstract: Landslide susceptibility evaluation can accurately predict the spatial distribution of potential landslides, which offers great usefulness for disaster prevention, disaster reduction, and land resource management. Aiming at the problems of insufficient samples for landslide compilation, difficulty in expanding landslide samples, and insufficient expression of nonlinear relationships among evaluation factors, this paper proposes a new evaluation method of landslide susceptibility combining deep autoencoder and multi-scale residual network (DAE-MRCNN). In the first step, a deep autoencoder network was used to learn the feature expression of the original landslide data in order to acquire effective features in the data. Next, a multi-scale residual network was constructed; specifically, the model was improved into a deep residual network model by adding skip connections in the convolutional layer. In addition, the multi-scale idea was utilized to fully extract the scale characteristics of the evaluation factors. Finally, the model was used for feature training, and the results were input into the Softmax classifier to complete the prediction of landslide susceptibility. For this purpose, a machine learning method and two state-of-the-art deep learning methods, namely SVM, CPCNN-ML, and 2D-CNN, were utilized to model landslide susceptibility in Hanzhong City, Shaanxi Province. The proposed method produced the highest model performance of 0.891, followed by 0.842, 0.869, and 0.873. The experimental results show that the DAE-MRCNN method can fully express the complex nonlinear relationships among the evaluation factors, alleviate the problem of insufficient samples in convolutional neural networks (CNN) training, and significantly improve the accuracy of susceptibility prediction.

Keywords: landslide susceptibility; deep autoencoder; convolutional neural networks; residual learning; multiscale

Citation: Wang, Z.; Xu, S.; Liu, J.; Wang, Y.; Ma, X.; Jiang, T.; He, X.; Han, Z. A Combination of Deep Autoencoder and Multi-Scale Residual Network for Landslide Susceptibility Evaluation. *Remote Sens.* **2023**, *15*, 653. https://doi.org/10.3390/rs15030653

Academic Editors: Yi Wang, Jun Hu and Weile Li

Received: 6 January 2023
Accepted: 19 January 2023
Published: 22 January 2023

1. Introduction

In recent years, the deteriorating global environment and increasing frequency of natural hazards have had a considerable impact on human production and living. In particular, landslides represent a major type of natural hazard [1,2]. Thus, landslide susceptibility evaluation has become the focus of disaster prevention and mitigation work; this challenging topic has attracted extensive attention from scholars at both domestic and international levels [3]. Landslide susceptibility evaluation, a method for predicting the temporal-spatial distribution and occurrence probability of landslide hazards, yields prediction results that can provide a vital scientific basis for landslide risk management, territorial spatial planning, and landslide monitoring [4]. Therefore, in order to enrich landslide susceptibility evaluation theories and models, protect people's lives and properties, reduce disaster losses, and improve the efficiency of disaster prevention and mitigation work, the study of landslide susceptibility evaluation and zoning offers academic significance while also providing a fact-based resource to help government departments formulate disaster prevention and mitigation measures, which has essential practical application value.

According to the different theoretical foundations, landslide susceptibility evaluation methods can be divided into two categories: deterministic and uncertainty-based [5,6]. The deterministic method primarily predicts the probability of landslide occurrence within a specific range by exploiting traditional quantitative physical–mechanical models based on landslide-inducing mechanisms and slope stability calculations. Therefore, this method is suitable for landslide susceptibility evaluation at both large scale and small range [7]. By comparison, the uncertainty method, which is further divided into two analysis methods, knowledge-driven and data-driven, superimposes weights on the factors affecting the occurrence of disasters, mainly by statistically analyzing historical disaster information at the locations where landslides have occurred and subsequently arriving at a prediction of the probability of landslide occurrence in the study area [8,9]. Since the uncertainty method considers the weight information of each indicator factor and has a high evaluation accuracy, it has found wide acclamation from scholars and researchers. Among the data-driven machine learning methods are those based on logistic regression (LR) [10,11], decision tree [12,13], random forest (RF) [14,15], and support vector machine (SVM) [16,17], all of which are well-suited to exploring the nonlinear relationship between landslide susceptibility and each evaluation factor. These methods are computationally efficient without much prior knowledge. However, the continuously increasing quality and quantity of geological hazard data is requiring the machine learning method to process a large amount of prior knowledge while also imposing various limitations because of the difficulty in efficiently obtaining the nonlinear relationships among evaluation factors from these complex data [18,19].

In contrast, deep learning has been recognized as better at converting low-level feature representations into deeper-level feature representations through multi-layer neural network training to further explore the distributed features of data; hence, this method has been widely used in landslide susceptibility mapping in recent years [20,21]. As one of the most representative methods in deep learning, convolutional neural networks (CNN) can better simulate the formation of landslide hazards with their powerful expressive learning capability to accurately predict potential landslide risk [22,23]. Wang et al. [24] applied CNN models to landslide susceptibility evaluation work by constructing CNN-1D, CNN-2D, and CNN-3D models for feature extraction and classification. Fang et al. [25] used a one-dimensional convolutional neural network (1D-CNN) with a five-layer network structure to evaluate landslide hazard susceptibility in Yongxin County, Jiangxi Province, and achieved good evaluation results. Youssef et al. [26] first used 13 evaluation factors as the input data set, then built a two-dimensional convolutional neural network model with a four-layer structure, and finally predicted the susceptibility index of each raster cell by the model to obtain the final landslide susceptibility results. These previous studies all achieved some success in improving classification accuracy. However, the current shortage of landslide inventories means difficulties in providing sufficient training samples for CNN. In addition, these single-scale network structures tend to neglect the details of evaluation factors and cannot effectively express the differences among attribute values.

Unlike the supervised CNN, autoencoder (AE) is an unsupervised feature learning neural network that can automatically learn from a large number of unlabeled data samples to acquire the effective features embedded within the data. This technique has been successfully applied to image processing [27,28], pattern recognition [29,30], anomaly detection [31], and so on. Considering the characteristics of AE and the deficiency that the single-scale network structure cannot effectively express the differences among attribute values, this paper proposes a landslide susceptibility evaluation method that integrates deep autoencoder and multi-scale residual networks. The new method makes full use of the data reconstruction advantage of AE along with the classification advantage of CNN networks. The proposed approach uses the organic integration of these two network structures to construct a unified architecture for susceptibility evaluation that can solve the problem of insufficient training samples and makes adequate use of the scale features of evaluation factors to extract more comprehensive discriminative features, significantly improving the accuracy of landslide susceptibility evaluation.

2. Study Area and Service Data

2.1. Study Area

China, located in the eastern part of the Asian continent, is a country with a high occurrence of geological disasters, including landslides, whose seriousness is underlined by their high frequency, wide influence, and persistence. For example, the severity of the problem of landslides is evident in a report published by China's Ministry of Natural Resources in 2021 that listed 4772 geological disasters in China, of which 2335 (48.93%) were landslides. Hanzhong City in Shaanxi Province features active geological tectonic movement, complex geological conditions, deep valleys, crushed terrain, and steep slopes that are affected by various factors such as engineering construction, land resource utilization, mineral resource development, and heavy rainfall, leading to increased landslide potential and making this location a key area for landslide prevention and control. Therefore, monitoring and forecasting landslide disasters in Hanzhong City, Shaanxi Province, is of immense practical significance.

Hanzhong City is in the southwest of Shaanxi Province, bounded by the Qinling Mountains in the north and the Daba Mountains in the south, covering a total area of 27,246 km^2. The center of Hanzhong Plain is formed by the alluvial accumulation of the Hanjiang River, which has a subtropical climate. The city is located in the middle region of the Hanzhong Basin, upstream of the Hanshui River Basin, between 105°30′50″ and 108°16′45″ E and 32°08′54″ and 33°53′16″ N. The city crosses the Qinling fold system and the Yangtze quasi-platform. Frequent tectonic movements and intense magmatic activities in the area's geological history complicate the regional geological situation. The terrain is generally low in the south and high in the north, and the city is comprised of three types of landforms: plains, hills, and mountains. The plain area is mainly distributed between 500 m and 600 m above sea level, with flat terrain and fertile soil, accounting for 34.62% of the city's area. Meanwhile, hilly areas are distributed between 601 m and 800 m above sea level, with large topographic relief, accounting for 28.1% of the city's area. Shallow mountains and middle mountains formed on the southern slope of the Qinling Mountains are the main mountainous areas in Hanzhong City, with a complex terrain. The altitude of this area ranges from 701 m to 2038 m, accounting for 37.2% of the city's total area. These complex geographic and geomorphic conditions jointly contribute to the natural environment of frequent landslide disasters in Hanzhong City. Hence, Hanzhong City was selected as the research area for this study.

2.2. Usage Data

The data on landslide points in Hanzhong City were derived from the Spatial Distribution Data of Geological Hazard Points from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences, acquired in 2017. The geological lithology data were obtained from the Spatial Distribution Data of Geological Lithology in China from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences, acquired in 2018. The rainfall data were derived from the China Annual Precipitation Spatial Interpolation Data Set Since 1980 from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences, acquired in 2018. The vegetation index data came from the Resource and Environmental Science Data Center of the Chinese Academy of Sciences, which generated the annual vegetation index data set with the maximum synthetic method, acquired in 2018. Digital elevation model (DEM) data and residential area data were obtained from the results of the first national geoinformation survey, acquired in 2018. The road data and water system data are public datasets provided by OpenStreetMap, acquired in 2018. In order to achieve the purpose of statistical analysis, the center of the practical range of landslide occurrence was adopted as the sampling point, and grid cells with a resolution of 30 m$\times$ 30 m were used as reference cells for division in this study. The study area was divided into 10,581 columns and 7769 rows, totaling 42,978,531 grid units.

3. Method

The procedure adopted for the proposed evaluation method to determine landslide susceptibility is illustrated in Figure 1. The constructed network is composed of three parts: selection of landslide susceptibility evaluation factors, construction of a training sample dataset, and susceptibility prediction based on DAE-MRCNN.

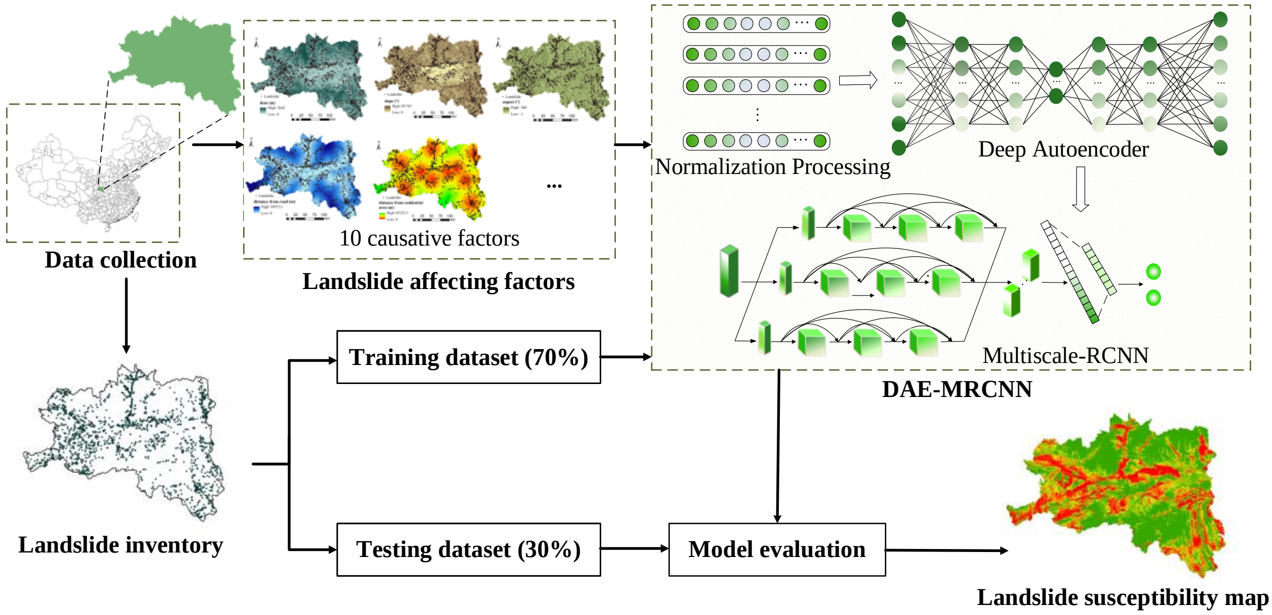

Figure 1. Flowchart of the proposed method.

3.1. Landslide Susceptibility Evaluation Factor

Taking into account the characteristics of the geological environment of landslide development in the study area, we conducted state grading and correlation analysis for each evaluation factor based on the landslide proportion, grading proportion, and information value under each indicator factor of landslide hazard. The formation of landslides is the result of the joint action of factors from various aspects; in addition, environmental variations in different areas make the selected evaluation factors for landslide susceptibility evaluation distinct in each area. For some evaluation factors with strong correlations, the correlations between the index factors could be eliminated to some extent by selective retention and elimination. Thus, the factors with a high frequency of occurrence were selected by referring to the general experience of landslide hazard research; meanwhile, the evaluated factors that strongly correlated with multiple index factors were eliminated. We began building the evaluation factor system of landslide susceptibility in the study area (Figure 2) by identifying the following 10 evaluation factors in 4 categories: topography (DEM, slope, aspect, and geomorphologic units landslide point density), human influence (distance from road and distance from the residential area), meteorological hydrology (cumulative rainfall and distance from river system), and geological and environmental factors (lithologic units landslide point density and normalized difference vegetation index [NDVI]).

Otherwise, considering the complicated geological structure research district, for discrete data such as lithology and geomorphology, combining the distribution of geological lithology space and geomorphological types and the distribution of landslide hazard points in Hanzhong City, the density of hazard points in each geological lithology unit and the hazard points in geomorphological units are quantified respectively.

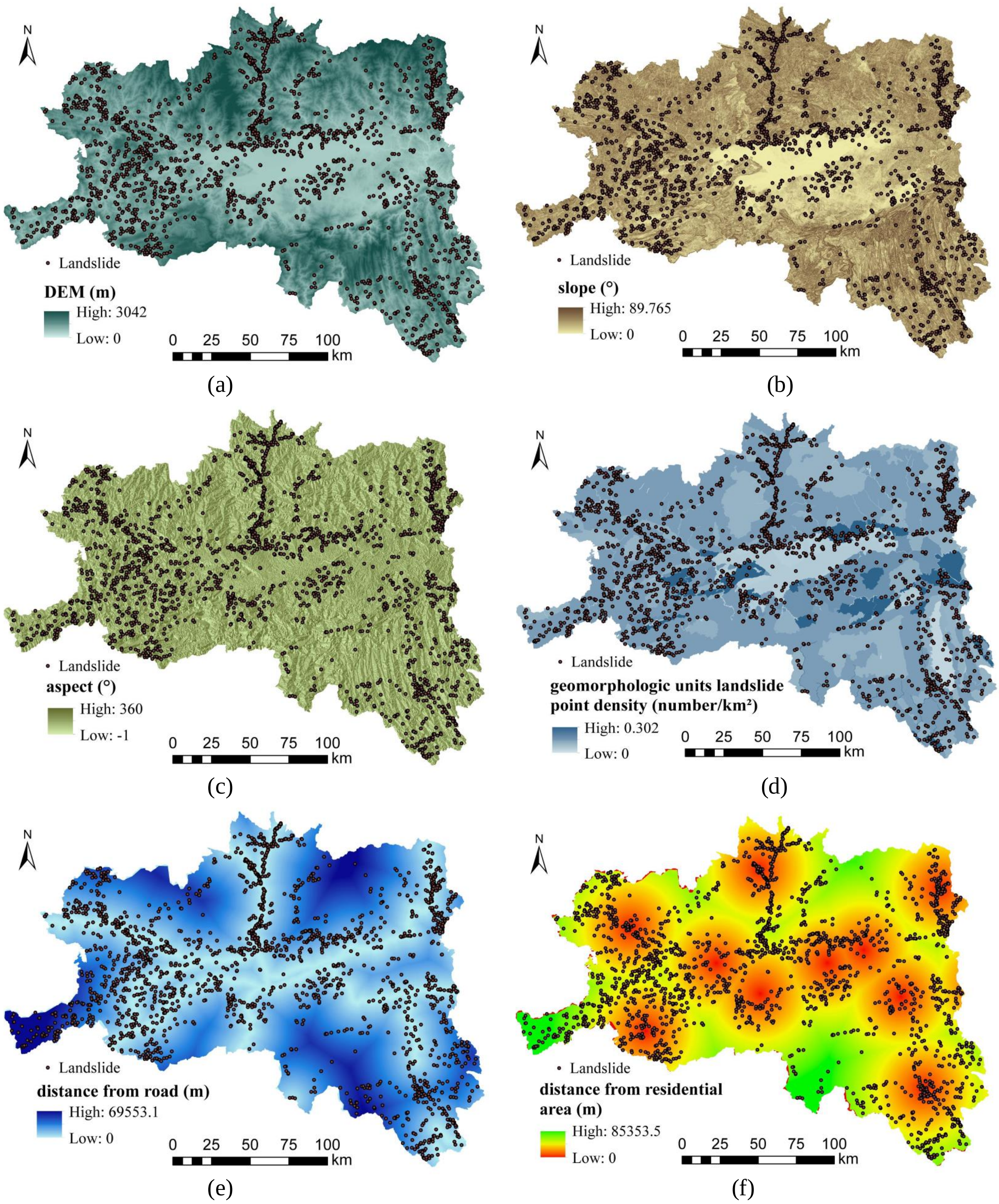

Figure 2. *Cont.*

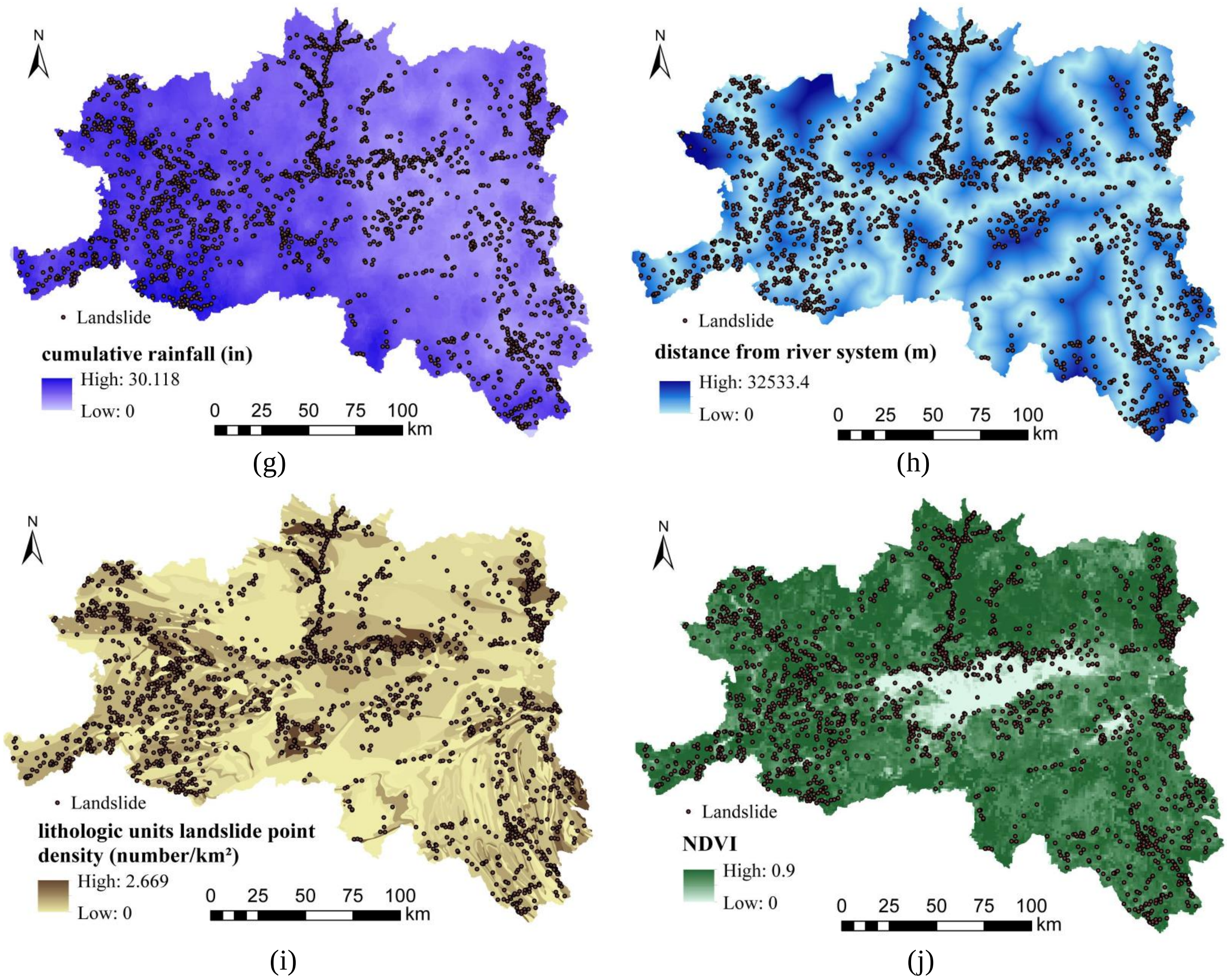

Figure 2. (**a**) DEM, (**b**) slope, (**c**) aspect, (**d**) geomorphologic units landslide point density, (**e**) distance from road, (**f**) distance from residential area, (**g**) cumulative rainfall, (**h**) distance from river system, (**i**) lithologic units landslide point density, (**j**) NDVI.

3.2. Evaluation Model Training Sample Dataset

Training the landslide susceptibility model required first establishing the landslide susceptibility evaluation training sample dataset. Each sample of data consisted of landslide attributes (landslide hazard unit or non-landslide hazard unit) and landslide evaluation factors (10 attribute values of landslide susceptibility evaluation factors), which were the model parameters of the evaluation model; the landslide attribute was the prediction target of the evaluation model.

In the process of selecting non-landslide points, we sought to avoid selecting samples comprising a raster unit with potential landslides by keeping the probability of landslides in the generated raster unit as low as possible. Specifically, we adopted the information value as the range constraint of the generation range of non-landslide points as follows: After grading each evaluation factor, the information value corresponding to the 10 evaluation factors was obtained and superimposed to obtain the total information map. Then, the very low susceptibility area and low susceptibility area, which were delineated based on the information value, were used as the generation areas of non-landslide samples. However, since there were still a few landslide points in the generation area, the selected non-landslide points were random and diverse due to distance constraints. On the basis of a thorough analysis of the landslide hazard development pattern in Hanzhong City, the distance between the generated non-landslide points and the existing landslide points, we defined the distance among the non-landslide points as not less than 1km as the distance constraint after repeated experiments.

Regarding the sample balance, when establishing the training sample dataset of the evaluation model, 2155 landslide points in Hanzhong City corresponded to 2155 raster cells. In the event that the above constraints were satisfied, the number of raster units consistent with the number of landslide points was randomly selected as non-landslide hazard units from the remaining 42,976,376 raster units. These two parts were defined as the sample dataset, 70% of which was used as the training dataset, and the other 30% was assigned to the testing dataset. In addition, a ten-fold cross-validation method was performed to determine the parameters of the model in the training dataset.

3.3. Landslide Susceptibility Prediction Based on DAE-MRCNN

The overall process of the DAE-MRCNN model constructed in this paper is shown in Figure 3. The model begins by normalizing the attribute values of the 10 evaluation factors. The next step involved a deep autoencoder network that was designed to achieve feature reconstruction, and then the reconstructed features were input into a depth residual module with three different scales for deep feature extraction. In the step that followed, the Softmax classifier was adopted to classify and obtain the affiliation degree. Finally, the affiliation degree was divided into five hazard susceptibility levels (very low susceptibility area, low susceptibility area, moderate susceptibility area, high susceptibility area, and very high susceptibility area) to generate the landslide susceptibility map, which could intuitively perceive the susceptibility degree of a landslide in each region of the study area.

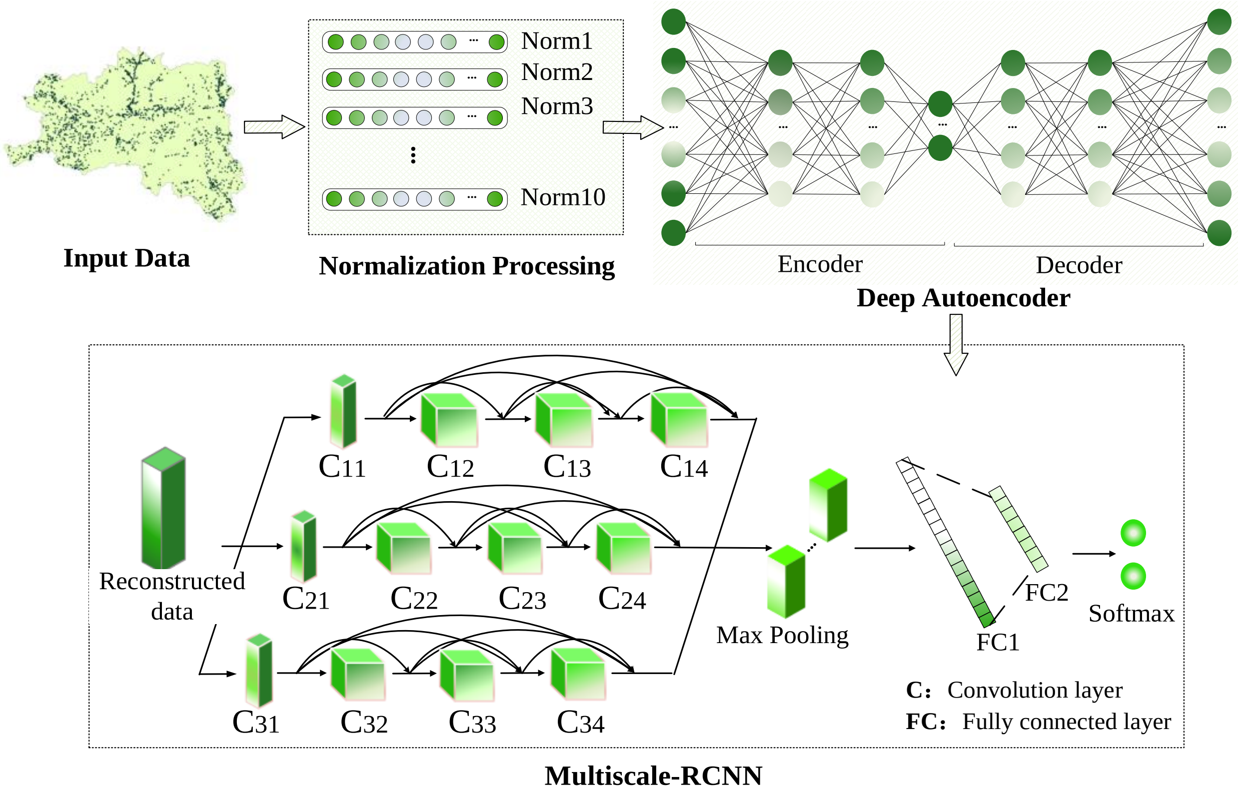

Figure 3. The overall flow of the DAE-MRCNN.

3.3.1. Normalize the Evaluation Factors

Because a certain variability, in terms of magnitude, of different evaluation factors existed, it was necessary to normalize each attribute value of evaluation factors to ensure the accuracy and validity of the susceptibility evaluation results. Without normalization, the convergence speed is slow, and the result obtained may be locally optimal. The normalization process is calculated as follows:

$$X = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \tag{1}$$

where Y is the input evaluation factor attribute value, X is the normalized evaluation factor attribute value, and $Y_{\max}$ and $Y_{\min}$ are the respective maximum and minimum values of the attribute values in the corresponding evaluation factors. After the normalization operation, the attribute values in each evaluation factor are in the range of [0, 1]; in other words, they have the same degree of perception.

3.3.2. Feature Reconstruction Based on DAE

Automatically learning the feature representation of raw data is one of the core purposes of deep learning and neural networks. AE is a nonlinear unsupervised learning network with strong feature learning ability that is extensively utilized in classification and unsupervised feature learning. It is well-suited to obtain the reconstruction information of the input data by characterizing the original data [32].

To acquire a better feature description than the original landslide data, we designed a deep autoencoder (DAE) network structure as illustrated in Figure 4. The DAE is composed of stacked connections of AE networks; each layer of the AE is trained separately, and the output of the previous AE layer is used as the input for the next layer of the AE, which has a better ability to extract data features. During the training process, the network parameters are constantly adjusted to minimize the objective function and achieve the purpose of reconstructing the input data in the decoding layer. The model yields a good reconstruction performance while improving the network generalization ability.

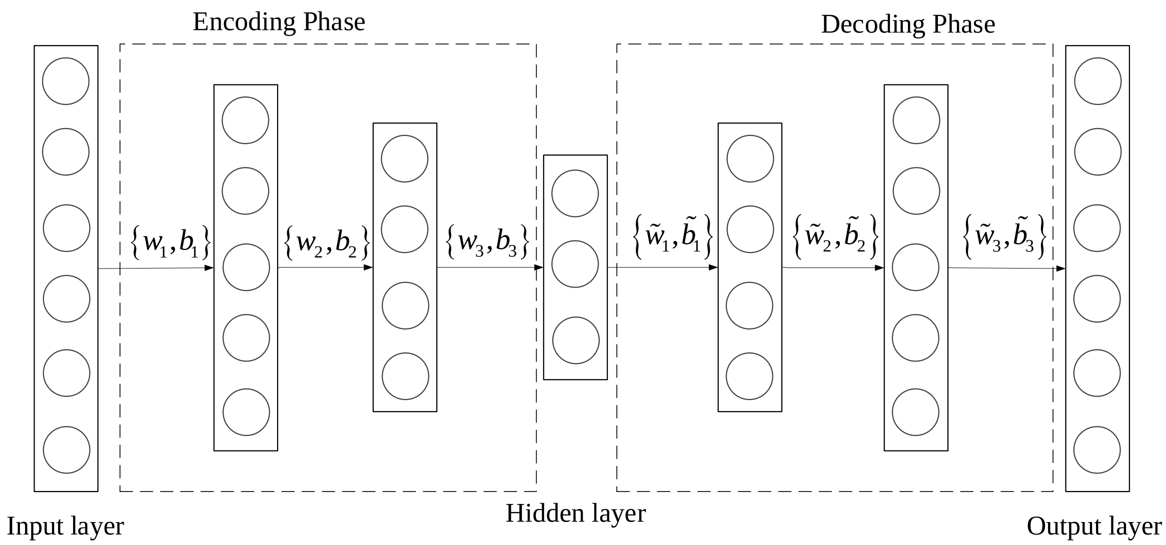

Figure 4. The structure of DAE.

Assuming that the normalized processed attribute value is $X = [x_{(1)}, x_{(2)}, \cdots, x_{(d_x)}]^{\mathrm{T}}$ $(x \in R^{d_x})$, where d_x is the dimension of the input data. In the encoding stage, x is mapped from the input layer to the hidden layer as $H = [h_{(1)}, h_{(2)}, \cdots, h_{(d_h)}]^{\mathrm{T}}(h \in R^{d_h})$, and d_h is the dimension of the hidden layer vector. In the decoding stage, h is mapped from the hidden layer to the output layer as $\widetilde{X} = [\widetilde{x}_{(1)}, \widetilde{x}_{(2)}, \cdots, \widetilde{x}_{(d_x)}]^{\mathrm{T}}(\widetilde{X} \in R^{d_x})$. In both stages, the transformation process of the input data is as follows:

$$h = \delta_f(WX + b) \tag{2}$$

$$\widetilde{X} = \delta_{\widetilde{f}}(\widetilde{W}H + \widetilde{b}) \tag{3}$$

where $W \in d_h \times d_x$ and $b \in R^{d_h}$ are the encoding weights and biases, $\widetilde{W} \in d_x \times d_h$ and $\widetilde{b} \in R^{d_x}$ are the decoding weights and biases, and δ_f and $\delta_{\widetilde{f}}$ are the activation functions, of which the sigmoid function is generally selected. On this basis, the loss function is minimized by continuously training $\left\{W, \widetilde{W}\right\}$ and $\left\{b, \widetilde{b}\right\}$ so that the reconstructed output $\widetilde{X}$ is as close as possible to the original input X.

3.3.3. Feature Extraction Based on Multiscale-RCNN

Considering the inadequate utilization of the detail features by existing single-scale techniques, we designed a multi-scale residual network (Multiscale-RCNN) based on Res-1DCNN to comprehensively obtain the feature information, as shown in Figure 3. The Multiscale-RCNN model can successively and alternatively extract complex nonlinear features from the evaluation factors. Compared with a simple CNN, the Multiscale-RCNN method fully fuses the features of evaluation factors among different branches by introducing a Res-1DCNN model which improves the classification accuracy.

In Figure 3, C_{11}-C_{14}, C_{21}-C_{24}, and C_{31}-C_{34} represent ordinary convolutional layers with a step size of 1 and kernel size of 1, 3, and 5, respectively (C_{11}, C_{21}, and C_{31} have a valid convolution mode, and the rest use the same convolution mode). Max pooling represents the maximum pooling layer with a step size of 2 and a pooling window size of 3, and FC represents a fully connected layer. First, the reconstructed landslide features are used as the input of the network. Next, a Res-1DCNN with a three-branch parallel structure (the convolutional kernel sizes of the three branches are 1, 3, and 5) is constructed, and the extraction of different perceptual fields is achieved through multiple parallel Res-1DCNN branches, which are utilized to connect the feature information extracted by each convolutional layer in the network. Following that step, the model extracts the deep features and then performs multi-scale feature fusion. Finally, a max pooling layer is introduced after the Multiscale-RCNN model to reduce the computational effort; the feature map output from the max pooling layer is converted into one-dimensional vectors and then connected to fully connected layers FC_1 and FC_2 with 512 and 256 neurons, respectively, to convert the feature data into 1×256 dimensional vectors.

Increasing the network depth of CNN has received much recognition in terms of better results in mining data features. However, this technique also tends to lead to the disappearance of the model learning gradient, resulting in a more difficult learning process. To solve this problem, we introduced a residual network (called ResNet [33]) into the model, which has previously been shown to achieve great success in the field of computer vision. The basic idea of residual learning is to stack multiple convolutional layers together as a block and then use a skip connection scheme to transfer the feature information from the previous layer to the next layer. This scheme allows fitting the residual mapping instead of the original identity mapping as a way to maintain the input information and broaden the propagated gradient. Figure 5 displays the ResNet constructed in this paper, where E is the input to the model, and $F(E)$ is the output after the linear transformation and activation of the layer, known as the residual.

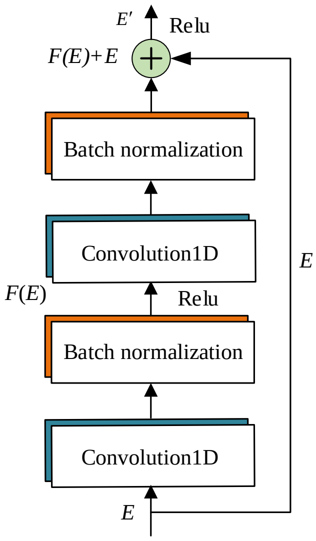

Figure 5. Residual Networks.

In the skip connection, we utilize convolution and batch normalization to extract sample feature E to get E'; the residual connection way is $F(E) + E$, and $F(E)$ is calculated as follows:

$$F(E) = \psi(W_2(\text{RelU}(\psi(W_1 + q_1))) + q_2) \tag{4}$$

where W_1 and W_2 are weights for each convolutional layer, q_1 and q_2 are biases, and ψ is represented as batch normalization.

The activation function can be used to enhance the nonlinear mapping ability of the network, and since the ReLU activation function converges faster than other activation functions, we utilize ReLU as the activation function of the network for all convolutional layers, which is expressed as:

$$\text{ReLU}(q) = max(0, q) \tag{5}$$

To enhance the reuse rate of feature information in CNN to a greater extent, as well as control the complexity of the model as much as possible, we constructed a deep residual model, Res-1DCNN, as shown in Figure 6 (C_1-C_4, R_1-R_6 represent the convolutional layers with a step size of 1, respectively). The network consists of convolutional layers, batch normalization layers, and dropout layers. Compared with the traditional residual unit structure, the Res1DCNN used in this paper makes adequate utilization of the correlation features among different layers in the model on the basis of keeping the number of parameters constant. This structure solves the problem of gradient disappearance due to the convolutional layers deepening, and extends the width of the network model on the basis of controlling the complexity. It can more effectively mine deeper features in the sample data and provide an effective network model for landslide susceptibility analysis.

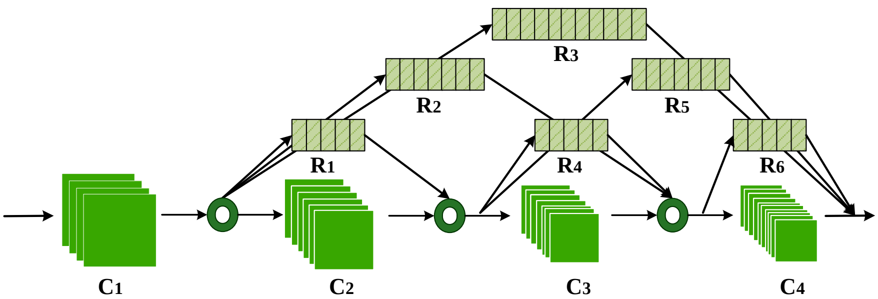

Figure 6. Res-1DCNN network structure.

In the construction of Res-1DCNN, according to the general experience of CNN design in computer vision, the complex structure of the evaluation factor means that the higher the convolution layer, the more convolution kernels are needed. In addition, in the network structure of residual learning, all convolutional layers use the same number of convolutional kernels to ensure that their dimensionality is consistent. Therefore, the number of convolutional kernels in the first convolutional layer C_1 is set to 8; next, C_2, C_3, and C_4 are set to 16, 32, and 64, respectively, and the step size of the convolutional kernels in C_1 to C_4 and R_1 to R_6 is set to 1. Otherwise, in order to speed up the training of the convolutional neural network and reduce the sensitivity to the network initialization, we adopt a batch normalization layer to normalize each input channel in batches [34].

As the number of network layers increased, the deep neural networks gradually produced a large number of parameters that could have easily led to an overfitting phenomenon. In addition, the slow learning speed of the large network structure poses great challenges to the overfitting problem. Dropout, a regularization method for neural network models, can randomly discard neurons and their connections from the neural network during the training process, reduce parameters in the network layer, and greatly alleviate the overfitting phenomenon [35,36]. Therefore, after fully connecting layers FC_1 and FC_2, we adopt a dropout regularization method to randomly discard a certain percentage of nodes in order to avoid the risk of overfitting. The dropout layer abandons part of the neurons with probability $p_{dropout}$, while the remaining neurons are retained with probability $1 - p_{dropout}$, and the output of the discarded neurons is set to 0, as shown in Figure 7.

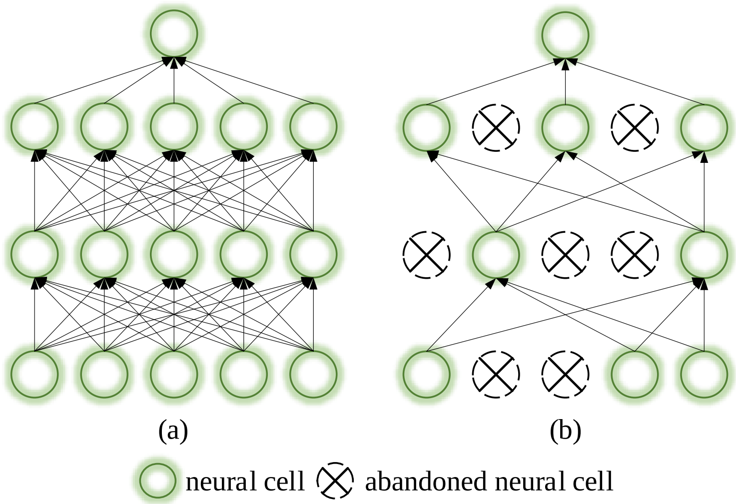

Figure 7. Dropout network structure. (**a**) Standard network, (**b**) Dropout network.

3.3.4. Loss Function

In the output layer of the network, the Softmax linear classifier is used for logistic regression classification, and the category corresponding to the maximum probability value is selected as the final prediction result to derive the subordinate degree. In practical terms, the cross-entropy loss function is often used in classification problems and, in particular, is frequently adopted as the loss function in the neural network training process. This function can be used to measure the similarity between the true category distribution and the model-predicted category distribution, referring to the model prediction to obtain the category probability and the one-hot encoding of the real category to calculate the cross-entropy loss function. The calculation formula for cross-entropy is shown below:

$$p_{ij} = \text{softmax}(\log(s_{ij})) = \frac{e^{\log(s_{ij})}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} e^{\log(s_{ij})}} \tag{6}$$

$$loss = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij} \ln p_{ij} \tag{7}$$

where n denotes the number of samples, k is the number of categories, y_{ij} represents the label of sample i, p_{ij} indicates the probability that sample i is predicted to be class j, s_{ij} is the score of class j in sample i, and $loss$ is the cross entropy of the true value and the predicted value.

4. Experimental Analysis

Various evaluation metrics were adopted to assess the performance of the proposed method, including precision, recall, F-measure, overall accuracy (OA), and AUC. Assuming that true positives (TP) and false positives (FP) were the number of landslide samples with correct and incorrect classification, respectively, false negatives (FN) and true negatives (TN) were the number of non-landslide samples with correct and incorrect classification, respectively.

Precision was the percentage of true landslide samples among all samples predicted as landslides.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall referred to the percentage of all landslide samples being accurately predicted as landslides.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

F-measure (F1) was the harmonic mean of precision and recall.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Overall accuracy (OA) was the comprehensive evaluation index of model predictions including landslide samples and non-landslide samples.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

AUC was the area under the receiver operating characteristic (ROC) curve, and the closer the value was to 1, the better the classification effect of the method was.

To ensure the reliability and accuracy of the experimental results, we took the average of the classification accuracy of 10 repeated experiments under each sample size as the final experimental result. In addition, Sections 4.2 and 4.3 present our comparative analysis in three aspects, under the condition that both training samples and test samples are the same: landslide susceptibility map, landslide susceptibility zoning statistics, and evaluation model accuracy. In terms of landslide susceptibility zoning, we considered various grading criteria to grade the evaluation factors according to the characteristics of each. Specifically, the natural discontinuity method is a commonly used classification method that can maximize the difference among classes. The geometrical interval method creates class breaks based on class intervals that have a geometric series; in addition, the coefficient in this classifier can be changed once (to its inverse) to optimize the class ranges. The geometrical interval method ensures that the interval of each class is relatively consistent, and it is often used when the data distribution is extremely skewed. When using the SVM method as a comparison experiment, the common natural discontinuity method was adopted for subordinate degree division. In addition, as a result of multiple feature extractions, the susceptibility evaluation method using deep learning generally has a higher recognition accuracy. In view of these characteristics, the geometric interval method was used as the division method in this study.

4.1. Parameter Settings

We divided the influencing factors of the proposed method into two parts: the deep autoencoder structure parameter and the multi-scale residual network parameter in feature extraction. The deep autoencoder structure parameter included the number of network layers and the number of output layer nodes. The parameters of the multi-scale residual network were the dropout probability value and the number of network iterations epoch. We first set Dropout=0.5 and Epoch=300 to analyze the influence of deep autoencoder network structure parameters on the model performance. After determining the optimal network structure, the effect of different values of the other two parameters on the model performance was considered, and the best parameters were selected.

4.1.1. Effect of the Number of Network Layers

The number of network layers in the deep autoencoder network played a critical role in the model performance, affecting the feature expression ability of the data. As the number of network layers increased, the feature expression ability was enhanced, and more abstract features could be obtained. However, an excessive number of layers could lead to overfitting of the model. In this experiment, six DAE-MRCNN network structures with different layers were set up, namely DAE^1-MRCNN, DAE^2-MRCNN, DAE^3-MRCNN, DAE^4-MRCNN, and DAE^5-MRCNN, and the corresponding network structures were represented as 256-10, 256-128-10, 256-128-64-10, 256-128-64-32-10, and 256-128-64-32-16-10, respectively. The influence of the number of network layers on the classification results is shown in Table 1. It is evident that the highest classification accuracy was achieved when the DAE-MRCNN adopted a network structure of four layers.

Table 1. Influence of the number of network layers on the classification results.

Network Structure	OA (%)	Precision	Recall	F1 Score
DAE^1-MRCNN	79.02	0.83	0.75	0.78
DAE^2-MRCNN	81.42	0.84	0.76	0.80
DAE^3-MRCNN	83.96	0.83	0.84	0.83
DAE^4-MRCNN	84.13	0.85	0.84	0.84
DAE^5-MRCNN	82.02	0.78	0.87	0.82

4.1.2. Effect of the Number of Output Layer Nodes

The number of the output layer nodes of the network model could express the ability of low-dimensional embedding of the original landslide data. Too small of a node number for output layer W led to insufficient feature expression ability, increasing the possibility of overfitting. Since the dimension of the original landslide data was 10, W was selected in [1,10]; its influence on the model performance is illustrated in Figure 8. It can be seen that the model performance improved as the number of output layer nodes increased; conversely, it appears to have stagnated or even decreased after gradually reaching the maximum value. This outcome can be explained by the fact that the increase in the number of output layer nodes was associated with the increasingly rich extraction of feature information. For the DAE^4-MRCNN method, the OA value was optimal when the number of output layer nodes was 9. At this point, as the feature dimension continued to increase, redundant information was gradually introduced into the network, which led to a deterioration in classification performance.

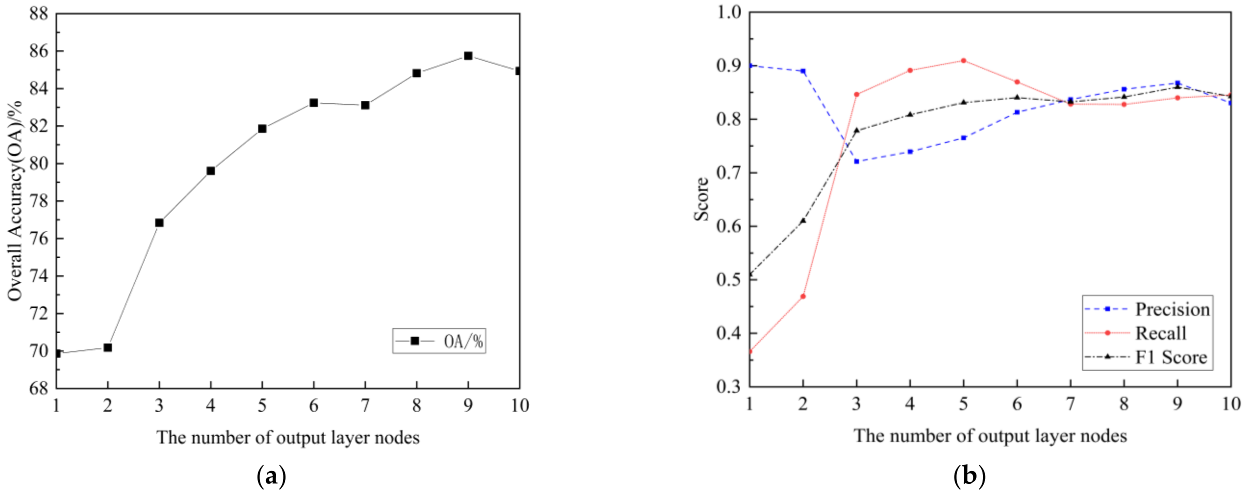

Figure 8. Influence of the number of output layer nodes on the classification result. (**a**) OA, (**b**) score.

4.1.3. Effect of Dropout Probability Value

The dropout layer prevented the model from relying too much on some local features by randomly discarding the parameter values of some neurons. Thus, its inclusion could enhance the robustness of the model while avoiding the overfitting phenomenon. The effect of different dropout values on the model performance is shown in Figure 9. It can be seen that the classification accuracy performance was optimal when the dropout value was 0.6.

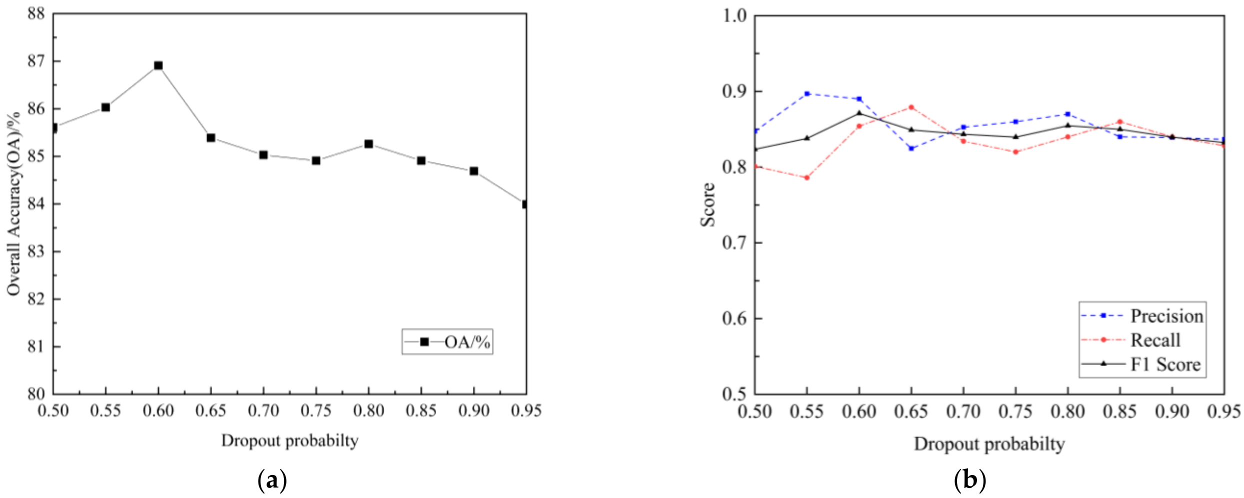

Figure 9. Classification performance of different dropout values. (**a**) OA, (**b**) score.

4.1.4. Effect of Iteration Times

During the process of network training, it is generally necessary to set an appropriate number of iterations to update the weights continuously in order to improve the generalization ability of the network. Figure 10 presents the influence of the different iteration times on the classification result. When the number of iterations is within 500, the classification accuracy has been maintained at about 86%. However, the loss value in the model decreased sharply, indicating that the model was still updating. Once the number of iterations reached 600, the loss value rapidly decreased from 0.11 to 0.02 and the score was at its highest. It then decreased slowly and gradually smoothed out, and the classification performance reached the optimum.

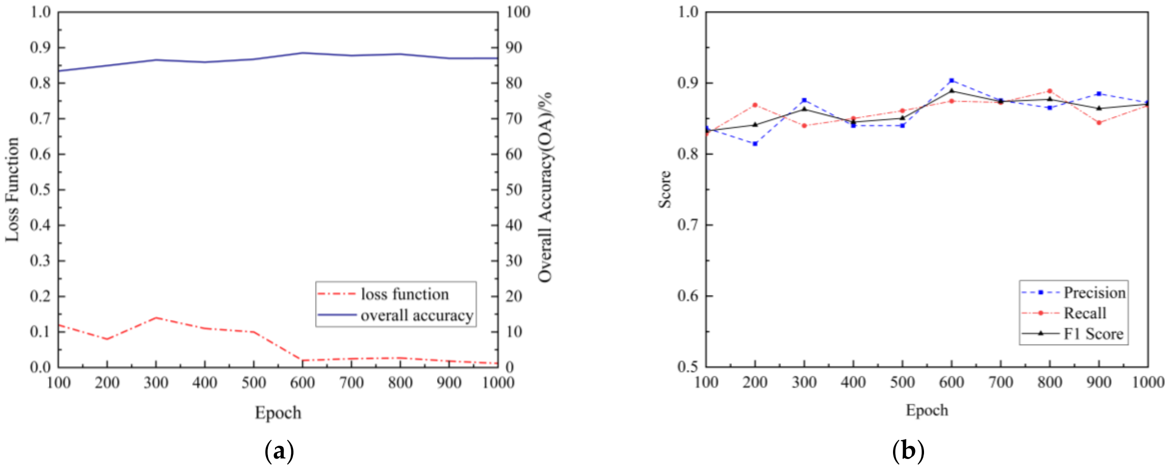

Figure 10. Influence of the different iteration times on the classification result. (**a**) Loss function and OA, (**b**) score.

4.2. Ablation Experiments

To verify the effectiveness of each model using the DAE-MRCNN method, we conducted ablation experiments for the models. The one-dimensional convolutional neural network designed in this study was chosen as the benchmark. The sub-models were defined as follows:

(1) The benchmark model (1DCNN) designed in this paper.
(2) The deep residual network formed by adding a residual learning model to the benchmark model (Res-1DCNN).
(3) The classification model consisting of the deep autoencoder network and the residual learning model added to the benchmark model (DAE-RCNN).
(4) The classification model consisting of the deep autoencoder network and the multi-scale deep residual network with convolution kernel sizes of 1, 3, and 5, added to the benchmark model, which was the classification method proposed in this study (DAE-MRCNN).

The maps of landslide hazard susceptibility corresponding to the proposed method and other sub-models are depicted in Figure 11. As a whole, the landslide susceptibility zoning obtained by the four models exhibited a certain similarity. The vast majority of the areas were very low susceptibility areas and low susceptibility areas. High susceptibility areas and very high susceptibility areas were mainly concentrated in the central part of Hanzhong City and extended in strips along the east and west directions, primarily along roads and rivers. Most of the moderate susceptibility areas were distributed in strips around the higher susceptibility areas, and a few were distributed around the lower susceptibility areas.

Taking the southern part of Hanzhong City as an example, the landslide susceptibility result obtained by the 1DCNN model was more scattered and less regular, which can be attributed to the fact that with limited training samples, although the 1DCNN model could extract features beneficial to the susceptibility evaluation, too many network layers might have caused the problem of overfitting. The division result of the Res-1DCNN model was more explicit than that of the 1DCNN model; this outcome resulted because the residual learning model could better overcome the problem of overfitting and improve the classification results to a certain extent. The DAE-RCNN model and the DAE-MRCNN model raised the hazard susceptibility by one level in one part of the area with the distribution of hazard points. This result illustrates that the deep autoencoder network was able to gain a better feature description than the original data while improving the generalization ability of the network, along with further extracted features that would be beneficial for landslide hazard susceptibility prediction. Compared with the DAE-RCNN model, the DAE-MRCNN model had relatively few discretely distributed high susceptibility areas, suggesting that the single-scale network structure tended to ignore the detail features of evaluation factors; in contrast, the multi-scale idea proposed in this study was able to extract more comprehensive discriminative features, effectively express the differences among attribute values, and improve the accuracy of landslide hazard susceptibility evaluation.

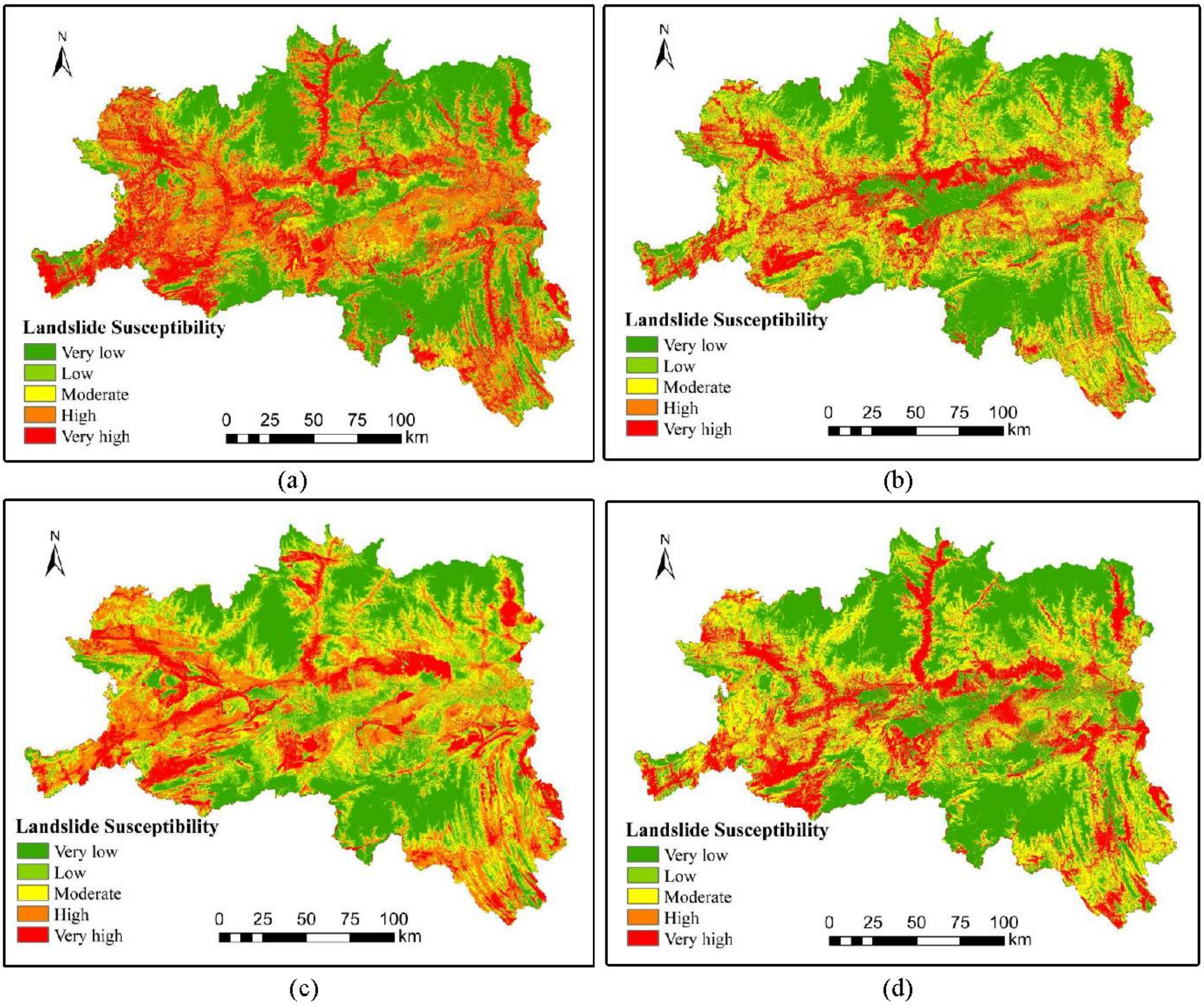

Figure 11. Landslide susceptibility maps of different sub-models. (**a**) 1DCNN, (**b**) Res-1DCNN, (**c**) DAE-RCNN, (**d**) DAE-MRCNN.

The statistical table of landslide hazard susceptibility zoning for the proposed method and other sub-models is displayed in Table 2. Figure 12 presents the ROC curves of different sub-models; in particular, the AUC values of 1DCNN, Res-1DCNN, DAE-RCNN, and DAE-MRCNN were 0.841, 0.858, 0.864, and 0.891, respectively. As the degree of hazard susceptibility increased, the density of landslide hazard points in each susceptible area also gradually increased, reaching the maximum in the very high susceptibility area, which is consistent with the actual situation. The classification performance of the 1DCNN model was relatively low because of the simple convolutional network structure adopted model, laying a certain foundation for the subsequent models. Taking high and very high susceptibility areas as examples, the class-specific accuracy of Res-1DCNN, DAE-RCNN, and DAE-MRCNN models was higher in each case than that of 1DCNN because the residual learning model in this study was more able to fully obtain the relevant features among different layers. In the DAE-RCNNN and DAE-MRCNN models, the sum of the percentages of these two zones was 88.81% and 88.86%, respectively, higher than the results for the other two methods, and the sum of the percentages in the low and very low susceptibility areas was 2.78% and 1.90%, respectively, lower than the results yielded by the other two methods. These outcomes point to the ability of the deep autoencoder network to accurately express the feature data and extract the features that are conducive to landslide hazard susceptibility evaluation, which further illustrates the effectiveness of each model of the proposed method.

Table 2. Statistical table of landslide susceptibility zoning for different sub-models.

Model	Class	Area	Landslides	Class-Specific Accuracy
1DCNN	Very low	12,912.25	31	0.002
	Low	3337.61	34	0.010
	Moderate	4935.76	131	0.026
	High	9402.41	586	0.062
	Very high	8092.63	1373	0.169
Res-1DCNN	Very low	12,580.65	25	0.002
	Low	4573.15	25	0.005
	Moderate	10,988.91	325	0.029
	High	3727.82	454	0.121
	Very high	6810.05	1326	0.194
DAE-RCNN	Very low	9508.76	10	0.001
	Low	6936.77	50	0.007
	Moderate	7343.59	181	0.024
	High	9967.52	817	0.081
	Very high	4924.01	1097	0.222
DAE-MRCNN	Very low	13,945.31	18	0.001
	Low	4552.32	23	0.005
	Moderate	9623.84	199	0.020
	High	3611.04	454	0.125
	Very high	6948.14	1461	0.210

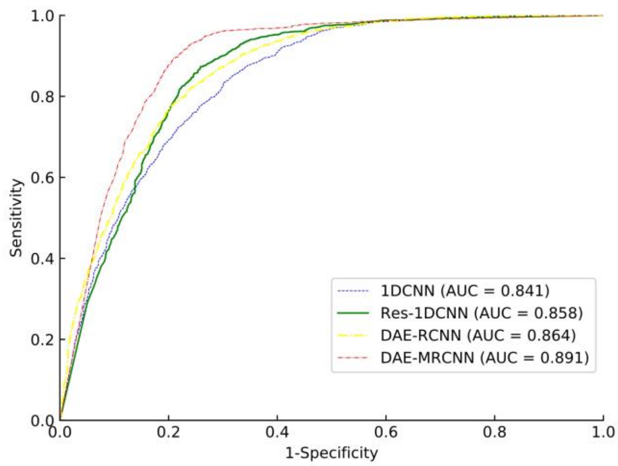

Figure 12. ROC curves of different sub-models.

4.3. Analysis of Landslide Susceptibility Zoning Results

To further verify the accuracy and reliability of the proposed method, we applied different methods and compared the predictions in four experiments.

(1) The method of directly classifying the original landslide data using an SVM classifier, and the radial basis function was adopted as the kernel function (SVM).
(2) The ensemble method based on a channel-expanded pre-trained CNN and a traditional machine learning model (CPCNN-ML) [23].

(3) The landslide susceptibility method based on the two-dimensional CNN method (2D-CNN) [37].
(4) The proposed method in the study.

The statistical table of landslide hazard susceptibility zoning for the different methods is shown in Table 3. The proposed method exhibited the lowest class-specific accuracy in the very low susceptibility and low susceptibility zones, and the highest class-specific accuracy in the very high susceptibility and high susceptibility zones, respectively. Figures 13 and 14 illustrate the landslide susceptibility maps and ROC curves of the four methods, respectively. The AUC values of the methods of SVM, CPCNN-ML, and 2D-CNN and the proposed method were 0.842, 0.869, 0.873, and 0.891, respectively. Compared with the other three methods, the SVM method divided more landslide points into the moderate susceptibility zone, and had the highest point density in the zones, indicating that the CNN method can better express the nonlinear relationships among evaluation factors. Although the AUC value of the CPCNN-ML method was improved to a certain extent, some zones with sparse distribution of landslide points were misclassified as very high susceptibility zone. The performance of zone delineation was similar between the 2D-CNN method and the DAE-RCNN method, but the DAE-RCNN method was more explicit in the delineation of landslide points. In summary, the proposed method eliminates the majority of zones directly from the risk zones in the practical application, which further demonstrates the proposed method is more reasonable and reliable.

Table 3. Statistical table of landslide susceptibility zoning for different methods.

Model	Class	Area	Landslides	Class-Specific Accuracy
SVM	Very low	11,206.88	37	0.003
	Low	8974.81	144	0.016
	Moderate	6824.08	235	0.034
	High	6162.37	679	0.110
	Very high	5512.51	1060	0.192
CPCNN-ML	Very low	9200.83	15	0.001
	Low	5042.66	22	0.004
	Moderate	5528.88	57	0.010
	High	10,595.48	460	0.043
	Very high	8312.80	1601	0.192
2D-CNN	Very low	11,530.76	18	0.001
	Low	5986.36	38	0.006
	Moderate	9207.62	256	0.027
	High	5840.31	562	0.096
	Very high	6115.60	1281	0.209
DAE-MRCNN	Very low	13,945.31	18	0.001
	Low	4552.32	23	0.005
	Moderate	9623.84	199	0.020
	High	3611.04	454	0.125
	Very high	6948.14	1461	0.210

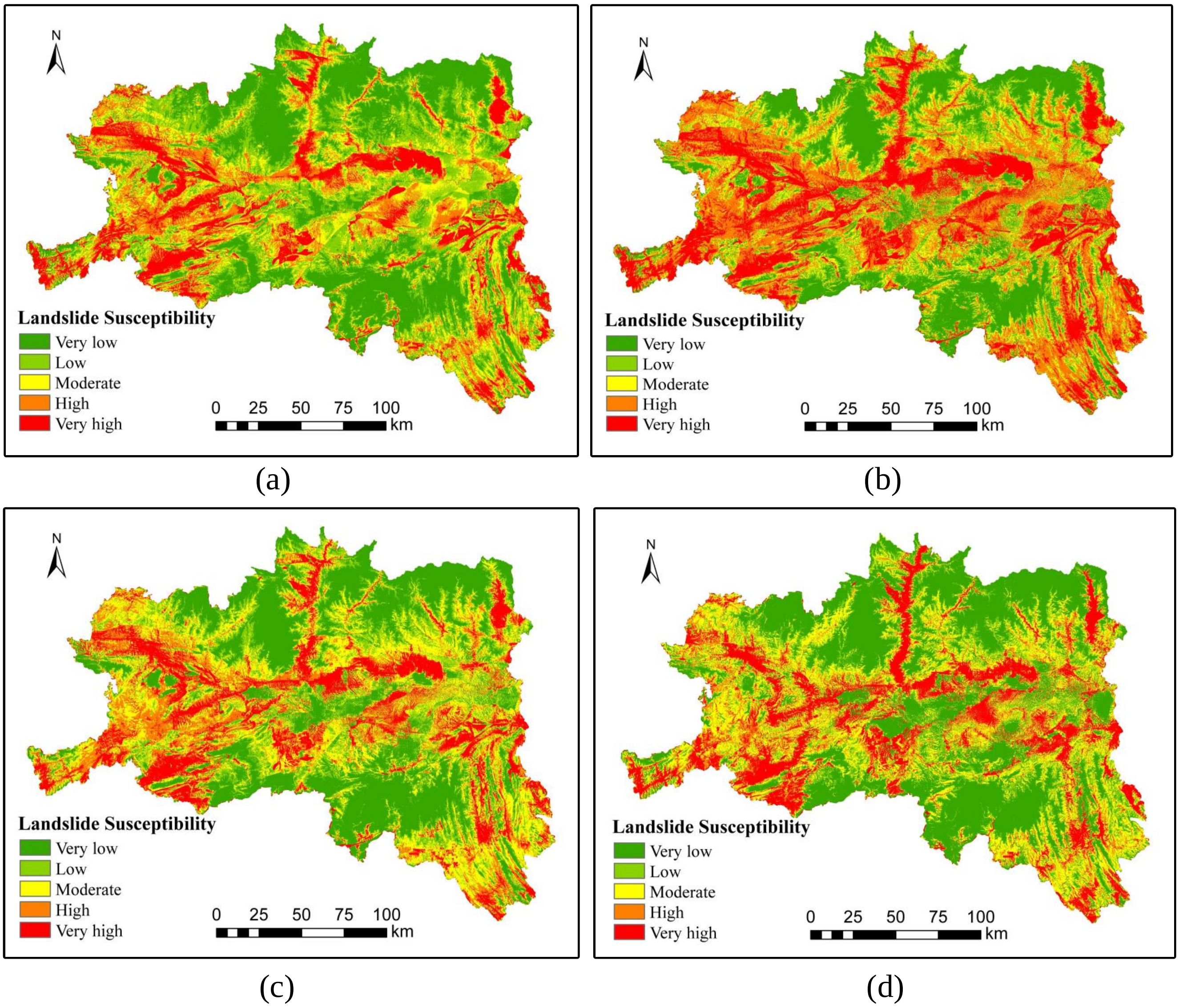

Figure 13. Landslide susceptibility maps of different models. (**a**) SVM, (**b**) CPCNN-ML, (**c**) 2D-CNN, (**d**) DAE-MRCNN.

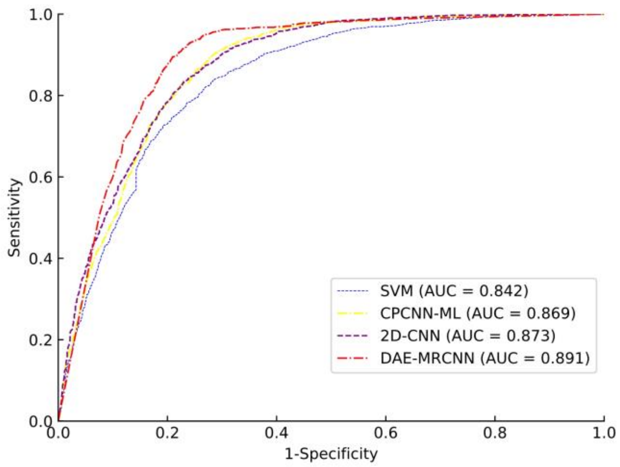

Figure 14. ROC curves of different models.

4.4. Discussion

Landslide susceptibility research provides a relatively accurate prediction of the spatial probability of a potential landslide occurrence in a certain region. Deep learning, a current significant branch of artificial intelligence, provides better simulation of landslide hazard formation and predicts potential risks effectively. In previous related literature, deep learning methods commonly suffer from insufficient samples of landslide, and gradient disappearance problems, and cannot effectively express complex nonlinear relationships among the evaluation factors. Therefore, we propose a new evaluation method of landslide susceptibility combining a deep autoencoder and a multi-scale residual network.

We conducted the experimental analysis of the major parameters affecting the performance of the model, including the number of network layers, the number of output layer nodes, the dropout probability value, and the number of network iterations epoch, so that the optimal model parameters were finally selected. For confirming the effectiveness of each model, the one-dimensional convolutional neural network designed in this study was chosen as the benchmark, and the ablation experiments for the models were conducted. These experimental results revealed that: (1) DAE was better able to express the complex nonlinear relationship among the evaluation factors in the case of insufficient landslide inventories; (2) the proposed depth residual model was caused by the increased number of network layers and adequately acquired the correlation features between different layers while keeping the number of parameters constant; and (3) the multi-scale idea made better use of the scale features of evaluation factors to extract more comprehensive discriminative features, providing the ability to solve the problem in which a single-scale network structure cannot effectively express the differences among attribute values.

To further verify the accuracy and reliability of the proposed method, a machine learning method and two state-of-the-art deep learning methods, namely SVM, CPCNN-ML, and 2D-CNN, were adopted as comparison experiments in this paper. The proposed method produced the highest model performance of 0.891, followed by 0.842, 0.869, and 0.873, while also achieving the highest class-specific accuracy in the high and very high susceptibility zones, and the lowest class-specific accuracy in the low and very low susceptibility zones. Therefore, the DAE-MRCNN method can effectively improve prediction performance and may be a promising method for future studies. Despite the excellent feature extraction capability of the DAE-MRCNN method, it is necessary to manually adjust a large number of parameters to optimize the complex CNN structure, which affects the efficiency of landslide hazard prediction and also makes it difficult to accurately obtain the optimal parameters of the model.

5. Conclusions

In this paper, we have presented a DAE-MRCNN method for landslide hazard susceptibility evaluation. Based on 2155 landslide hazard sites in Hanzhong City, 10 influencing factors, such as slope, aspect, and DEM, were selected as landslide hazard susceptibility evaluation factors, and the DAE-MRCNN method was designed to evaluate landslide hazard susceptibility in the study area.

In summary, the proposed method not only effectively improves the prediction accuracy of landslide susceptibility but also lays a firm foundation for the application of deep learning in the field of susceptibility evaluation. However, the large number of parameters involved in CNN makes it necessary to manually adjust the parameters for optimization. Solving these problems and investigating more effective classification methods will be the goal of future research.

Author Contributions: Conceptualization, methodology, writing—original draft, software, validation, Z.W. and X.M.; supervision, writing—review and editing, funding acquisition, project administration, S.X.; funding acquisition, project administration, Y.W. and J.L.; formal analysis, data curation, T.J., X.H. and Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Key Research and Development Program of China (2020YFC1511704 and 2022YFC3005705).

Data Availability Statement: Data sharing is not applicable to this article.

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

References

1. Li, W.; Zhu, J.; Fu, L.; Zhu, Q.; Guo, Y.; Gong, Y. A rapid 3D reproduction system of dam-break floods constrained by post-disaster information. *Environ. Model. Softw.* **2021**, *139*, 104994. [CrossRef]
2. Li, W.; Zhu, J.; Fu, L.; Zhu, Q.; Xie, Y.; Hu, Y. An augmented representation method of debris flow scenes to improve public perception. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 1521–1544. [CrossRef]
3. Peng, T.; Chen, Y.; Chen, W. Landslide Susceptibility Modeling Using Remote Sensing Data and Random Subspace-Based Functional Tree Classifier. *Remote Sens.* **2022**, *14*, 4803. [CrossRef]
4. Maheshwari, B. Spatial predictive modelling of rainfall-and earthquake-induced landslide susceptibility in the Himalaya region of Uttarakhand, India. *Environ. Earth Sci.* **2022**, *81*, 1–24.
5. Westen, C.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [CrossRef]
6. Regmi, N.R.; Giardino, J.R.; Vitek, J.D. Assessing susceptibility to landslides: Using models to understand observed changes in slopes. *Geomorphology* **2010**, *122*, 25–38. [CrossRef]
7. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [CrossRef]
8. Chen, W.; Li, W.; Chai, H.; Hou, E.; Li, X.; Ding, X. GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji City, China. *Environ. Earth Sci.* **2016**, *75*, 1–14. [CrossRef]
9. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. 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* **2016**, *13*, 361–378. [CrossRef]
10. Wang, L.J.; Sawada, K.; Moriguchi, S. Landslide-susceptibility analysis using light detection and ranging-derived digital elevation models and logistic regression models: A case study in Mizunami City, Japan. *J. Appl. Remote Sens.* **2013**, *7*, 3561. [CrossRef]
11. Rai, D.K.; Xiong, D.; Zhao, W.; Zhao, D.; Zhang, B.; Dahal, N.M.; Wu, Y.; Baig, M.A. An Investigation of Landslide Susceptibility Using Logistic Regression and Statistical Index Methods in Dailekh District, Nepal. *Chin. Geogr. Sci.* **2022**, *32*, 834–851. [CrossRef]
12. Arabameri, A.; Chandra Pal, S.; Rezaie, F.; Chakrabortty, R.; Saha, A.; Blaschke, T.; Di Napoli, M.; Ghorbanzadeh, O.; Thi Ngo, P.T. Decision tree based ensemble machine learning approaches for landslide susceptibility mapping. *Geocarto Int.* **2022**, *37*, 4594–4627. [CrossRef]
13. Li, Y.; Lin, F.; Luo, X.; Zhu, S.; Li, J.; Xu, Z.; Liu, X.; Luo, S.; Huo, G.; Peng, L. Application of an ensemble learning model based on random subspace and a J48 decision tree for landslide susceptibility mapping: A case study for Qingchuan, Sichuan, China. *Environ. Earth Sci.* **2022**, *81*, 1–12. [CrossRef]
14. Liu, Y.; Zhang, W.; Zhang, Z.; Xu, Q.; Li, W. Risk Factor Detection and Landslide Susceptibility Mapping Using Geo-Detector and Random Forest Models: The 2018 Hokkaido Eastern Iburi Earthquake. *Remote Sens.* **2021**, *13*, 1157. [CrossRef]
15. Lin, R.; Liu, J.; Xu, S.; Liu, M.; Zhang, M.; Liang, E. Evaluation method of landslide susceptibility based on random forest weighted information. *Sci. Surv. Mapp.* **2020**, *45*, 131–138. [CrossRef]
16. Chang, Z.; Du, Z.; Zhang, F.; Huang, F.; Guo, Z. Landslide Susceptibility Prediction Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine Learning Models. *Remote Sens.* **2020**, *12*, 502. [CrossRef]
17. Xu, S.; Liu, J.; Wang, X.; Zhang, Y.; Lin, R.; Zhang, M.; Liu, M.; Jiang, T. Entropy index incorporated into support vector machine for landslide hazard susceptibility evaluation method a case of Shaanxi Province. *Geomat. Inf. Sci. Wuhan Univ.* **2020**, *45*, 1214–1222. [CrossRef]
18. Saha, S.; Saha, A.; Hembram, T.K.; Mandal, K.; Sarkar, R.; Bhardwaj, D. Prediction of spatial landslide susceptibility applying the novel ensembles of CNN, GLM and random forest in the Indian Himalayan region. *Stoch. Environ. Res. Risk Assess.* **2022**, 1–20. [CrossRef]
19. Hakim, W.L.; Rezaie, F.; Nur, A.S.; Panahi, M.; Khosravi, K.; Lee, C.W.; Lee, S. Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea. *J. Environ. Manag.* **2022**, *305*, 114367. [CrossRef]
20. Habumugisha, J.M.; Chen, N.; Rahman, M.; Islam, M.M.; Ahmad, H.; Elbeltagi, A.; Sharma, G.; Liza, S.N.; Dewan, A. Landslide susceptibility mapping with deep learning algorithms. *Sustainability* **2022**, *14*, 1734. [CrossRef]
21. Guo, C.; Wu, J.; Zhao, S.; Wang, Z.; Meena, S.R.; Zhang, F. Landslide susceptibility assessment based on multi GPUs: A deep learning approach. *CCF Trans. High Perform. Comput.* **2022**, *4*, 135–149. [CrossRef]
22. Machichi, M.A.; Saadane, A.; Guth, P.L. Subimages-Based Approach for Landslide Susceptibility Mapping Using Convolutional Neural Network. In *Geospatial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 131–141.
23. Chen, Y.; Ming, D.; Xiao, L.; Lv, X.; Zhou, C. Landslide Susceptibility Mapping Using Feature Fusion-Based CPCNN-ML in Lantau Island, Hong Kong. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3625–3639. [CrossRef]

24. Wang, Y.; Fang, Z.; Hong, H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Sci. Total Environ.* **2019**, *666*, 975–993. [CrossRef] [PubMed]
25. Fang, Z.; Wang, Y.; Ling, P.; Hong, H. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Comput. Geosci.* **2020**, *139*, 104470. [CrossRef]
26. Youssef, A.M.; Pradhan, B.; Dikshit, A.; Al-Katheeri, M.M.; Mahdi, A.M. Landslide susceptibility mapping using CNN-1D and 2D deep learning algorithms: Comparison of their performance at Asir Region, KSA. *Bull. Eng. Geol. Environ.* **2022**, *81*, 1–22. [CrossRef]
27. Patel, H.; Upla, K.P. A shallow network for hyperspectral image classification using an autoencoder with convolutional neural network. *Multimed. Tools Appl.* **2022**, *81*, 695–714. [CrossRef]
28. He, Y.; Carass, A.; Zuo, L.; Dewey, B.E.; Prince, J.L. Autoencoder Based Self-Supervised Test-Time Adaptation for Medical Image Analysis. *Med. Image Anal.* **2021**, *72*, 102136. [CrossRef]
29. Yu, J.; Zheng, X.; Wang, S. A deep autoencoder feature learning method for process pattern recognition. *J. Process Control* **2019**, *79*, 1–15. [CrossRef]
30. Wu, L.; Zhang, X.; Zhang, X.; Chen, X.; Chen, X. Metric learning for novel motion rejection in high-density myoelectric pattern recognition. *Knowl.-Based Syst.* **2021**, *227*, 107165. [CrossRef]
31. Zhang, C.; Hu, D.; Yang, T. Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost. *Reliab. Eng. Syst. Saf.* **2022**, *222*, 108445. [CrossRef]
32. Dolgikh, S. On Unsupervised Categorization in Deep Autoencoder Models. In Proceedings of the International Conference on Computer Science, Engineering and Education Applications, Kyiv, Ukraine, 21–22 February 2022; Springer: Cham, Switzerland, 2020; pp. 255–265.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
35. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
36. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *Comput. Sci.* **2012**, *3*, 212–223.
37. Renza, D.; Cárdenas, E.A.; Martinez, E.; Weber, S.S. CNN-Based Model for Landslide Susceptibility Assessment from Multispectral Data. *Appl. Sci.* **2022**, *12*, 8483. [CrossRef]