Landslide Susceptibility Prediction Modeling Based on Remote Sensing and a Novel Deep Learning Algorithm of a Cascade-Parallel Recurrent Neural Network

Li Zhu [1], Lianghao Huang [1], Linyu Fan [1], Jinsong Huang [2], Faming Huang [3,*], Jiawu Chen [3], Zihe Zhang [1] and Yuhao Wang [1]

[1] Information Engineering School, Nanchang University, Nanchang 330031, China; lizhu@ncu.edu.cn (L.Z.); wanfy@email.ncu.edu.cn (L.H.); 416114417186@email.ncu.edu.cn (L.F.); 411014419075@email.ncu.edu.cn (Z.Z.); wangyuhao@ncu.edu.cn (Y.W.)

[2] ARC Centre of Excellence for Geotechnical Science and Engineering, University of Newcastle, Newcastle, NSW 2308, Australia; jinsong.huang@newcastle.edu.au

[3] School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China; jiawu_chen@email.ncu.edu.cn

* Correspondence: faminghuang@ncu.edu.cn; Tel.: +86-1500-277-6908

check for updates

Abstract: Landslide susceptibility prediction (LSP) modeling is an important and challenging problem. Landslide features are generally uncorrelated or nonlinearly correlated, resulting in limited LSP performance when leveraging conventional machine learning models. In this study, a deep-learning-based model using the long short-term memory (LSTM) recurrent neural network and conditional random field (CRF) in cascade-parallel form was proposed for making LSPs based on remote sensing (RS) images and a geographic information system (GIS). The RS images are the main data sources of landslide-related environmental factors, and a GIS is used to analyze, store, and display spatial big data. The cascade-parallel LSTM-CRF consists of frequency ratio values of environmental factors in the input layers, cascade-parallel LSTM for feature extraction in the hidden layers, and cascade-parallel full connection for classification and CRF for landslide/non-landslide state modeling in the output layers. The cascade-parallel form of LSTM can extract features from different layers and merge them into concrete features. The CRF is used to calculate the energy relationship between two grid points, and the extracted features are further smoothed and optimized. As a case study, the cascade-parallel LSTM-CRF was applied to Shicheng County of Jiangxi Province in China. A total of 2709 landslide grid cells were recorded and 2709 non-landslide grid cells were randomly selected from the study area. The results show that, compared with existing main traditional machine learning algorithms, such as multilayer perception, logistic regression, and decision tree, the proposed cascade-parallel LSTM-CRF had a higher landslide prediction rate (positive predictive rate: 72.44%, negative predictive rate: 80%, total predictive rate: 75.67%). In conclusion, the proposed cascade-parallel LSTM-CRF is a novel data-driven deep learning model that overcomes the limitations of traditional machine learning algorithms and achieves promising results for making LSPs.

Keywords: landslide susceptibility prediction; deep learning; cascade-parallel recurrent neural network; conditional random field; logistic regression; multilayer perceptron; decision tree; remote sensing; geographic information system

1. Introduction

Landslides are one of the most common geological disasters worldwide and cause considerable damage to public infrastructure and human life every year [1,2]. The predictive modeling of landslide occurrence is one of the main challenges in geological hazard research. A landslide susceptibility map (LSM) is an effective visualization technology for the localization of a landslide region and sustainable land management [3–5]. Moreover, the landslide susceptibility model based on geological environmental conditions can supply the government with an important theoretical basis for land resource planning and disaster prevention and reduction.

The process of landslide susceptibility prediction (LSP) modeling primarily includes a catalog of landslides, environmental factors extraction, model architecture construction, model training, landslide susceptibility mapping, and model evaluation [6,7]. The catalog of landslides (landslide area, boundary, locations) are measured using global positioning systems and put into a geographic information system (GIS) for landslide storage and management [8,9]. The environmental factors are extracted from the remote sensing (RS) images, such as Landsat8 TM image, digital elevation model (DEM), aerial imagery, and LiDAR, based on the GIS spatial analysis, including terrain analysis, hydrological analysis, and map algebra [10]. As a whole, the LSP modeling is built on the platform of GIS because of the spatial big data analysis, storage, mapping, and management abilities [11].

Importantly, based on the above obtained spatial data sources, the input data, network architecture, parameter settings, and optimization algorithm of the model all affect the accuracy of landslide predictions. In recent decades, researchers have developed various predictive models combined with the GIS, which includes heuristic models and statistical models [12], e.g., information value model [13], logistic regression [14,15], entropy index [16], certainty factor [17,18], analytic hierarchy process [19,20], etc. However, these models have certain limitations in LSP applications. For example, models often require feature data to be subject to a certain statistical distribution or an independent and identically distributed assumption, but environmental factors often fail to meet these prior knowledge requirements. In recent years, machine learning methods have been widely used in landslide susceptibility modeling and have achieved remarkable results because of their high prediction accuracy and absence of a prior knowledge requirement. As such, this approach produces a higher prediction accuracy, can more precisely identify the nonlinear relationship between input and output variables, and retains more characteristic information from the original data [21–24]. This approach includes multiple adaptive regression splines [25,26], fuzzy logic [27,28], artificial neural network [15,29], multilayer perceptron [30], decision tree [31–33], random forest [34–36], support vector machine [37–39], rule-based approach [40], and multi-criteria evaluation techniques [41], among others.

Selected disadvantages of traditional machine learning models occur in the application of LSP: (1) the models generally require a large amount of prior knowledge and assumptions; (2) the networks are not sufficiently deep to fully extract the underlying landslide features; (3) the networks are not sufficiently wide to consider the correlations between sub-regions; and (4) the models encounter problems, such as over-fitting, time-consuming computation, ease of falling into local optima, and sensitivity to missing data, which affect the accuracy of prediction.

Deep learning is an emerging multilayer neural network learning algorithm that can overcome the shortcomings of traditional machine learning models to a certain extent. Compared with traditional machine learning methods, deep-learning-based models are capable of extracting inherent and deep features. Deep learning methods are data-driven without requiring additional prior knowledge or assumptions [42–44], and can effectively identify useful information among miscellaneous data and obtain the optimal parameters for constructing models in the process of model training. At the same time, the impact of over-fitting on model prediction accuracy can be eliminated by using a considerable number of iterations. Moreover, as advanced data mining models, deep learning algorithms have been widely used in various fields, such as image recognition [44], face recognition [45], medical artificial intelligence [46], natural hazards mapping [47], etc. Due to their strong capability of feature extraction, it is necessary to apply deep learning methods to predict the landslide susceptibility in the study

area [48]. The current basic models of deep learning include the deep neural network, convolutional neural network, recurrent neural network, and a series of new structures, such as the long short-term memory (LSTM) [49] structure and residual network (ResNet) [50], which are state-of-the-art models with a high performance across numerous applications and are valuable in the theoretical study of deep learning, among others. In addition, many methods, such as the rectified linear unit activation function, have been developed to solve the problems of gradient disappearance and over-fitting in traditional networks. Moreover, model calibration is necessary for landslide susceptibility prediction. Venables and Ripley [51] present a guide to using S environments to perform statistical analyses, practical problems, and real data sets. Common calibration methods for landslide succeptibility (LS) assessment includes (i) linear discriminant analysis (LDA) [51–53], (ii) quadratic discriminant analysis (QDA) [53], (iii) logistic regression (LR) [54], and (iv) neural network (NN) modeling [53,55]. A comprehensive calibration for neural-network-based LSP modeling aims to improve the output characteristic of all nodes, resulting in a computationally intensive problem.

In summary, landslide susceptibility prediction plays a significant role in resisting landslides and reducing disasters. It is also a challenging problem because landslide features are generally uncorrelated or nonlinearly correlated, resulting in limited LSP performance when leveraging traditional machine learning models. To overcome the limitations where traditional machine learning algorithms require substantial prior knowledge and achieve promising results regarding LSPs, a novel data-driven deep learning model using the LSTM recurrent neural network and conditional random field (CRF) in cascade-parallel form was proposed for LSP based on the remote sensing (RS) images and a geographic information system (GIS). In the proposed model, the cascade-parallel LSTM can extract and merge features from different layers without prior knowledge, and CRF is used to further smooth and optimize those extracted features. Meanwhile, a dropout strategy is adopted to prevent the problem of over-fitting in traditional neural networks.

The study and comparison of machine learning models based on deep learning algorithms and other traditional models for LSP is of great significance. Taking Shicheng County in China as a case study, this study proposed a recurrent neural network, namely a cascade-parallel LSTM model, to predict landslide susceptibility in the study area. Considering the influence of the relationship between adjacent grids on landslide susceptibility, the conditional random field (CRF) was further introduced to optimize the models. Furthermore, logistic regression (LR), multilayer perceptron (MLP), and C5.0 decision tree (C5.0 DT) methods were selected for analysis and comparison.

2. Materials and Methods

2.1. Materials

2.1.1. Introduction to Shicheng County

Shicheng County is located in the northeast portion of Ganzhou City, Jiangxi Province, with a longitude ranging from 116°05′46″–116°38′03″ N, a latitude ranging from 25°57′47″–26°36′13″ E and a total area of 1581.53 km^2, as illustrated in Figure 1. Shicheng County is a typical southeast low mountain and hilly region, the topography of which is enveloped by many mountains in the northeast, rolling hills in the southwest, and flat terrain in the middle. Moreover, this region resides in a subtropical humid monsoon climate zone with ample sunshine, an annual mean temperature of 18.1 °C, and abundant rainfall, where the annual precipitation is 1919.6 mm. Furthermore, Shicheng basin formation in this area is attributed to the influence of tectogenesis, and the Guifeng Group and the Late Cretaceous of the Ganzhou Group are its main outcropping strata. The Guifeng Group is the prime landscape formation of the Danxia landform, which is divided into the Hekou formation with cracked landforms in its red sandstone, and the Tangbian and Lianhe formations, from bottom to top. The Anyuan-Yingtan fault zone, the Caledonian Indosinian-Huali granite, and the Cretaceous basin, which are distributed in the form of beads, are located on the west side of the basin. The Heyuan-Shaowu fault zone, which

has obvious control of the formation and distribution of the Meso Cenozoic basin, is located on the east side.
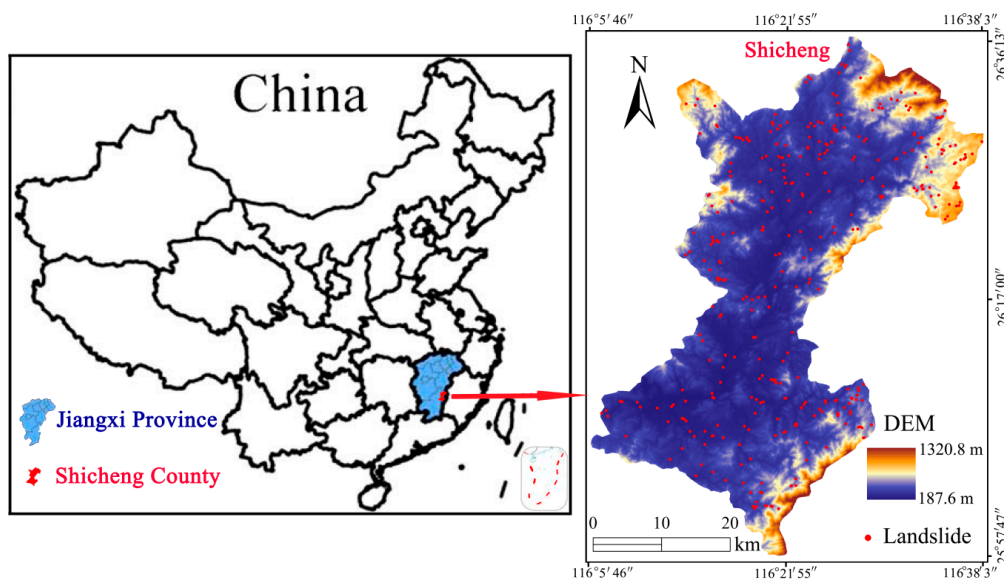

Figure 1. Location map in the study area. DEM: Digital elevation model.

2.1.2. Landslide Distribution Map

The catalog of the landslide distribution map is the key to the LSP, where the accurate landslide locations and detailed geological information are primarily reflected. The landslide locations are acquired based on historical landslide disaster reports, field investigations, and interviews with residents conducted by the Land and Resources Department of Jiangxi Province. A total of 369 landslides have been recorded in Shicheng County from 1970 to 2012 (Figure 1). The main features of these landslides can be described as small-scale, high-frequency, with a wide distribution.

The total area covered by the landslides in Shicheng County is about 2.44×10^6 m^2 with the smallest area being 1.0×10^3 m^2 and the largest area being 1.6×10^4 m^2. The landslide masses are generally composed of quaternary silty clay intercalated with crushed stones, and the depth of these sliding masses ranges from 2 m to 8 m. In addition, these landslides can be regarded as shallow soil landslides with a movement type of clay/silt slide [56]. Finally, the landslides in Shicheng County are mainly triggered by the seasonal heavy rainfall and unreasonable constructive activities (such as slope toe cutting and road network construction)

2.2. Environmental Factors Extraction from RS and GIS

A total of 14 environmental factors were extracted from RS and the GIS, including topographic, land cover, hydrological, and lithology factors.

2.2.1. Terrain Analysis Using DEM Data

The topographic factors (including elevation, slope, aspect, profile curvature, plan curvature, and relief amplitude) and hydrological factors (distances to rivers and topographic wetness index (TWI)) were extracted from the DEM with a 30-m resolution (http://gdem.ersdac.jspacesystems.or.jp). The profile curvature represents the vertical plane parallel to the slope direction; it is calculated as the slope of the slope using the terrain analysis tool in ARCGIS 10.2, ESRI, USA. Meanwhile, the plan curvature represents the various features of the concave terrain from the horizontal direction and is calculated as the slope of the aspect [57] in ARCGIS software. Meanwhile, the relief amplitude, which represents the slope surface relief characteristics of Shicheng County, was calculated using the maximum height difference method in ARCGIS software [58].

2.2.2. Land Cover Factors Acquisition from Google Earth and Landsat TM8 Images

The environmental factors of the total surface radiation and population density were downloaded from Google Earth 7.1.8.3036 (32-bit). The total surface radiation, defined as the sum of direct and diffuse solar radiation received on the horizontal surface, affects the hydrological environments and land cover types of the slope [59]. Population density, which represents the population in a certain area, mainly plays an important role in human construction activities [60].

The Landsat TM8 image taken on 5 October 2013, path/row 121/42 with a 30-m resolution (http://ids.ceode.ac.cn/index.aspx) was used to extract the land cover factors, which are expressed using the normalized difference vegetation index (*NDVI*), the modified normalized difference water index (*MNDWI*), and the normalized difference build-up index (*NDBI*) [61]. *NDVI* mainly represents the vegetation growth and cover rates of the study area (Equation (1)). *MNDWI* represents the surface water distribution features, as shown in Equation (2). In addition, the *NDBI* reflects the building distribution features on the surface of the landslide, as shown in Equation (3). The $P(B3)$, $P(B4)$, $P(B5)$, and $P(B6)$ represent the visible green band, visible red band, near-infrared band, and middle infrared band, respectively, of the Landsat 8 TM image.

$$NDVI = \frac{P(B5) - P(B4)}{P(B5) + P(B4)} \tag{1}$$

$$NDBI = \frac{P(B6) - P(B5)}{P(B6) + P(B5)} \tag{2}$$

$$MNDWI = \frac{P(B3) - P(B6)}{P(B3) + P(B6)} \tag{3}$$

2.2.3. Analysis of Hydrological Factors

The effects of hydrological factors on landslides are reflected through the topographic wetness index (*TWI*) and the distances to rivers. *TWI* represents the effects of topography and soil moisture content on the probability of landslide occurrence. The distances to rivers represents the distance of grid cells to the rivers and drainages in the area, suggesting the effects of surface water on landslides. The distances to rivers can be calculated based on the multi-ring buffering method in ARCGIS software.

The river networks of the study area are extracted through the hydrological analysis tool [62]. First of all, the DEM data was filled by the fill tool of the hydrological analysis tool to handle the defects of the original data. Second, the flow direction was calculated based on the filled DEM data, then the flow accumulation of each grid cell was calculated based on the flow direction and the filled DEM. Third, the flow accumulation threshold was set to 5000 according to the trial-and-error method [63], and the grid cells with a flow accumulation threshold greater than 5000 could form the river networks. Finally, a river network with a linear vector format could be obtained through the grid turn line tool in ARCGIS software.

TWI, representing the impacts of topographic factors on the soil moisture content along the runoff areas, has been widely used to describe the hydrological influences on landslide occurrences. *TWI* can be expressed as given in Equation (4), where A_s refers to the upstream catchment area and β represents the slope angle of a certain grid cell:

$$TWI = \ln(A_s / \tan \beta) \tag{4}$$

2.2.4. Analysis of the Lithology Factor

Lithology is represented by the rock types, which are the material basis of landslide development, with a great influence on the shear strength and permeability of the rocks and soils in the landslide mass. In this study, a rock type map was provided by the Land and Resources Department of Jiangxi Province and was mapped through the spatial analysis tool of ARCGIS software. The rock types in Shicheng County are mainly defined as metamorphic, carbonate, and clastic rocks.

2.3. Frequency Ratio (FR) Method

The application of the *FR* method is usually based on the assumption that future landslide events will occur in the geological environment of the past and present landslides [48,64]. Moreover, *FR* quantifies the relationship between the observed landslides and each environmental factor and is defined as the ratio of the proportion of landslide grid numbers in each subclass factor to the proportion of study area grid numbers in each subclass factor. The higher the *FR*, the higher the probability that a landslide will occur in the subcategory of the corresponding factors:

$$FR_{ja} = \frac{S'_{ja}/S}{M'_{ja}/M} \tag{5}$$

where FR_{ja} is the frequency ratio of the ath subcategory in the jth environmental factor, S_{ja}' is the number of landslide grids of the ath subcategory in the jth environmental factor, S is the total number of landslide grids, M'_{ja} is the number of study grids of the ath subcategory in the jth environmental factor, and M is the number of total grids in the study area.

2.4. Modeling Processes of Cascade-Parallel LSTM-CRF

The proposed cascade-parallel LSTM-CRF model for LSP consists of five steps:

(1) A large landslide inventory and related environmental factors are collected from a global positioning system, RS, and GIS technologies.
(2) The frequency ratio values (*FR*s) of those environmental factors are calculated and labeled, and then are used as input variables of the machine learning models.
(3) The landslide and non-landslide grid cells are used as the output variables of these models.
(4) The landslide susceptibility models are built based on the cascade-parallel LSTM-CRF, as well as the other models for comparisons.
(5) The LSMs of Shicheng County and the model accuracy evaluation are performed.

2.5. Theory of Cascade-Parallel LSTM-CRF

In this study, the cascade-parallel LSTM and CRF were proposed for LSP, as shown Figure 2. The model consists of sub-region feature modeling, fully connected layer classification, and implicit state modeling. The proposed model has a higher LSP performance and overcomes the limitations in terms of achieving a wider range of landslide data prediction. With the stacked structure, it can extract more comprehensive and accurate landslide features, which facilitates classification. Furthermore, CRF can optimize the extracted features and smooth the predicted results of mutations.
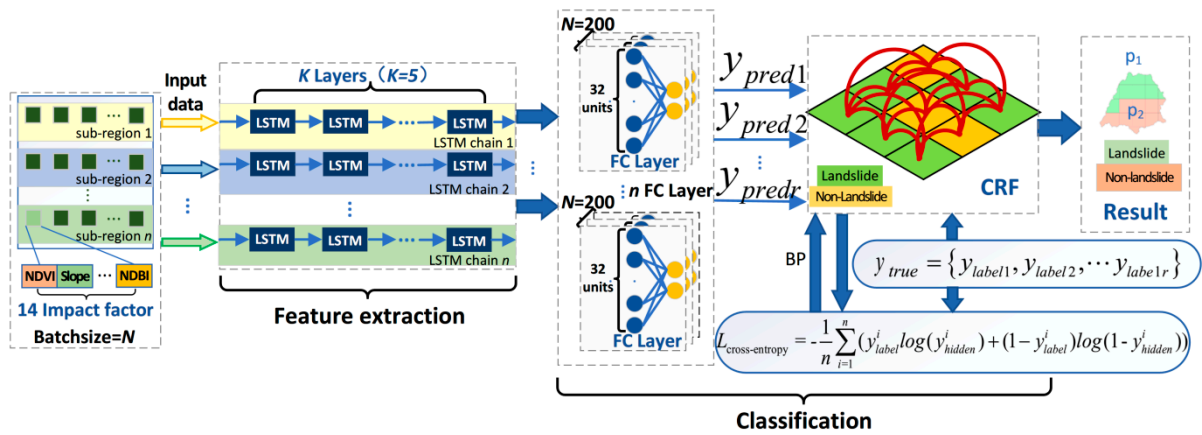
Figure 2. Algorithm flowchart. BP: Back propagation, CRF: Conditional random field, FC: Full connection, LSTM: Long short-term memory.

2.5.1. Sub-Region Feature Modeling

The modeling area is first divided into consecutive sub-regions S^e, ($e = 1, 2, \ldots, \lceil M/N \rceil$), where M is the number of grids in the whole area, N is the defined number of grids in each sub-region, and $\lceil \bullet \rceil$ denotes the ceiling operation (i.e., x is rounded up to the nearest integer). The *FR* vectors x^e of the J environmental factors in the eth sub-region are used as the raw input data of the eth cascade LSTM. As shown in Figure 2, the batch size is equal to the number of grids in each sub-region. The batch gradient descent (BGD) method is used in each sub-region. In light of the complicated characteristics of landslide factors, this section uses cascade LSTM [65] to extract the corresponding features of landslides, which shows superior performance over the single LSTM.

By leveraging the cascade of K LSTM cells and fully connected feed-forward units, the deep features and space patterns of the input *FR* vectors x^e are extracted for landslide/non-landslide classification. The LSTM cell is shown in Figure 3.
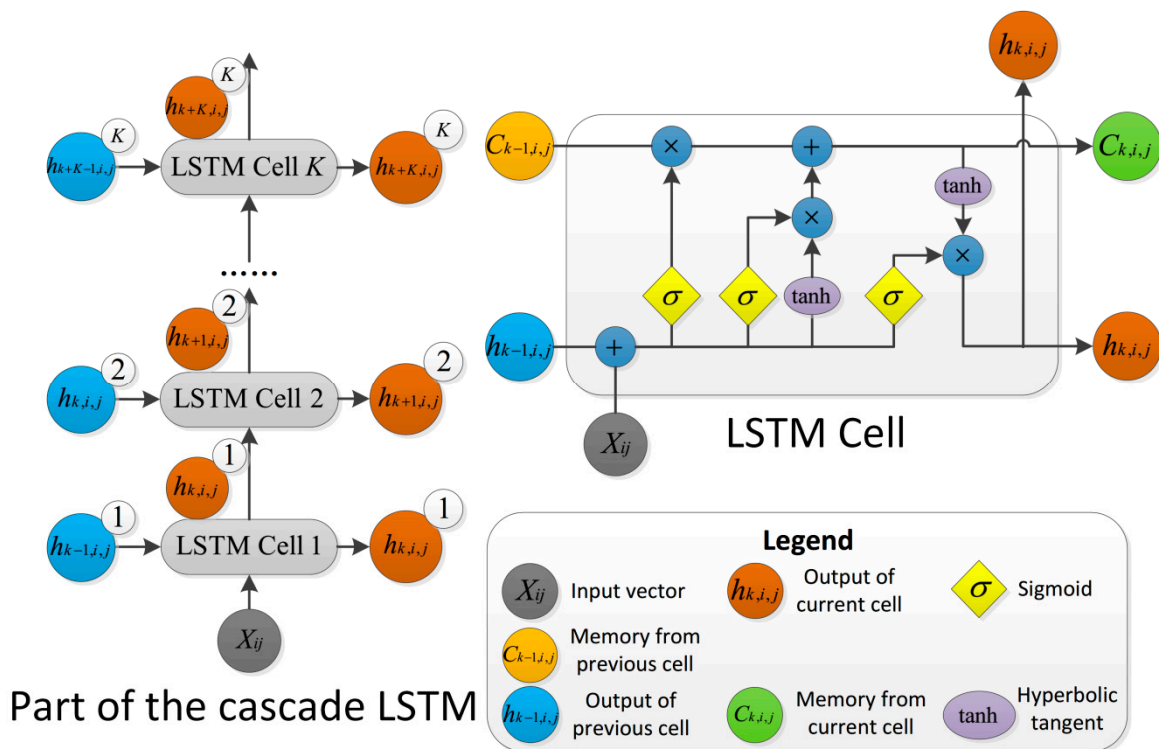

Figure 3. LSTM structure diagram.

Due to the high coupling and sophisticated nonlinear correlations among landslide environmental factors, landslide environmental factors cannot be used as features of prediction. Therefore, it is necessary to establish a cascade deep network structure that extracts effective features of different weights from different layers by decoupling the complex relationships through multiple implicit layers. The input layer of the cascade-parallel LSTM-CRF is the real matrix $R^{i \times j}$ of the landslide environment factors, where the row vector $\{i \,|1 \leq i \leq M, i \in Z\}$ is the number of grids and M is the total number of grids, the column vector $\{j \,|1 \leq j \leq J, j \in Z\}$ is the number of landslide environment factors and J is the total number of landslide environment factors, and $R^{M \times J}$ passes through an implicit layer of K LSTM cells cascaded in series. The LSTM unit is shown in Figure 3.

When the kth LSTM cell is processing the data of the jth environmental factor of the ith grid point using forward propagation, the update formulas for the forget gate layer ($f_{k,i,j}$), the input gate layer ($i_{k,i,j}$), and the output gate layer ($o_{k,i,j}$) are written as follows:

$$f_{k,i,j} = \sigma\left(W_f \cdot \left[h_{k-1}, x_{ij}\right] + b_f\right) \tag{6}$$

$$i_{k,i,j} = \sigma\left(W_i \cdot \left[h_{k-1}, x_{ij}\right] + b_i\right) \tag{7}$$

$$o_{k,i,j} = \sigma\left(W_o \cdot \left[h_{k-1}, x_{ij}\right] + b_o\right) \tag{8}$$

where "σ" represents the sigmoid function used as the activation function; W_f, W_i, and W_o are the weight matrices of each gate of the LSTM cell; h_{k-1} and x_{ij} represent the hidden state from the previous unit and the jth environmental factor of the ith grid point, respectively; b_f, b_i, and b_o are the corresponding bias terms; and $f_{k,i,j}$ outputs a number between 0 and 1. The internal data calculation process of LSTM is as follows:

$$\widetilde{C}_{k,i,j} = \tanh\left(W_c \cdot \left[h_{k-1}, x_{ij}\right] + b_C\right) \tag{9}$$

$$C_{k,i,j} = f_{k,i,j} * C_{k-1} + i_{k,i,j} * \widetilde{C}_{k,i,j} \tag{10}$$

$$h_{k,i,j} = o_{k,i,j} * \tanh(C_{k,i,j}) \tag{11}$$

where $C_{k,i,j}$ is the current cell state, $\widetilde{C}_{k,i,j}$ is the candidate state generated by the current input gate layer based on the previous hidden layer, and $h_{k,i,j}$ is the hidden layer state. First, the input gate layer generates the candidate state value $\widetilde{C}_{k,i,j}$. Second, the forget gate layer decides which information is discarded (when $f_{k,i,j} = 1$, it means that the previous cell state is completely retained; when $f_{k,i,j} = 0$, it means that the previous cell state is completely discarded). The value of the new candidate state is added to the input gate layer to create an update of the cell state. Finally, the output gate layer outputs $C_{k,i,j}$ and the hidden variable $h_{k,i,j}$ based on the current cell state. During the process of forward propagation, the formula for the change from the input of raw data to the output from cascade LSTM is as follows:

$$y_{out} = \prod_{k=1}^{K} \sum_{i=1}^{M} \sum_{j=1}^{J} o_{k,i,j} * \tanh\left(f_{k,i,j} * C_{k-1} + i_{k,i,j} * C_{k,i,j}\right) \tag{12}$$

where M, J and K represent the total number of grid points, the environmental factors of the single grid point, and the LSTM cells, respectively. The internal weights of the LSTM cells and bias parameters are updated using backpropagation. When the kth LSTM cell is processing the ith grid, the error of an output cell is as follows:

$$\delta_k(i) = \frac{\partial L_{cross-entropy}}{\partial x_i} \tag{13}$$

The error of the entire output gate layer can be readily computed based on the error of the output cell as follows:

$$\delta^o(i) = \sigma'(o(i)) \left(\sum_{v=1}^{K} \tanh(h_v(i)) \sum_k W_c^k \delta_k(i) \right) \tag{14}$$

The three gate layers of the LSTM act as three switches, which have the function of the selective forgetting of old cell states, selective logging of input information, and selective determination of which portion of the cell state is output. The three gates of LSTM [66] control the state of the cells, which are similar to the conveyor belt, making one-dimensional data transmit throughout the chain smoothly. Hence, LSTM is ideal for processing 1D (one-dimensional) landslide raw data. In this study, the algorithm can fuse the features of each layer of the LSTM through the data stream connection between layers, which extracts more abundant features than the single layer of LSTM.

To prevent the model from over-fitting during training, a dropout process was adopted in the LSTM cells. The specific process is described as follows. First, in a batch of training samples, a portion of neurons in the LSTM cells are randomly deleted, and the remaining neurons are fed to the next layer. Second, after obtaining this batch of training samples, the deleted neurons are restored, and certain neurons in the LSTM cells are randomly removed once again. The corresponding parameters are updated via the Adam method, which is performed on the un-removed neurons. This process is repeated. The dropout in this study was set to 0.2, which remarkably reduced the probability of over-fitting. This approach is beneficial to the subsequent independent and non-redundant optimal feature extraction. For the entire area, $\lceil M/N \rceil$ sub-regions are predicted with parallel implementation. The cascading LSTMs are applied to each sub-region.

2.5.2. Fully Connected Layer Classification

LSP is a two-category task. Therefore, a classification layer is required in the model to predict whether the grid has a landslide. The *K* cascade LSTM extracts the 32-dimensional landslide features of the landslide factors for each grid, and all features form a fully connected layer. A classification layer formed by two neurons follows the fully connected layer. This process maps the 32-dimensional feature space to the 2-dimensional space through a sigmoid function, i.e., the landslide/non-landslide space. Additionally, the nonlinear sigmoid function maps the real features in $(-\infty, +\infty)$ to the landslide probability values in $[0, 1]$, producing a landslide prediction.

2.5.3. Hidden State Modeling

The extracted features are observed states, and the landslide/non-landslide is the hidden state. With the fully connected output as the conditional random field input, the extracted features are further optimized and classified. For the LSP task (or general structured prediction), it is beneficial to consider the correlations between labels in neighborhoods, and it is reasonable to assume that the labels in each input sub-region are correlated.

The landslide factor is the observed state, and the landslide/non-landslide is the implied state. The prediction results obtained by the fully connected layer are all from the observed state, without consideration of the spatial correlation between the hidden states. For example, if the grids are spatially related and other raster predictions in a certain neighborhood of a certain grid are landslides, then the grid is more likely to contain a landslide. For example, based on the assumption that a spatial correlation exists between grids, if other grids in a certain neighborhood of a grid are predicted to be landslides, then this grid is more likely to contain a landslide. In this study, according to the order characteristics of landslide data, a CRF [67,68] is introduced after the fully connected layer for hidden state modeling to produce a landslide/non-landslide prediction sequence $y_{hidden} = \left\{ y_{hidden}^1, y_{hidden}^2, \cdots, y_{hidden}^n \right\}$ and make more accurate predictions. At the same time, CRF can smooth out the sudden changes in the prediction results and correct the unreasonable prediction. Assuming that the fully connected layer output $y_{hidden} = \left\{ y_{hidden}^1, y_{hidden}^2, \cdots, y_{hidden}^n \right\}$ and the true label sequence during the training process phase

$y_{label} = \{y_{label}^1, y_{label}^2, \cdots, y_{label}^n\}$ satisfy the Markov property, then the conditional probability $p\left(y_{lable}\middle|y_{pred}\right)$ is calculated as follows:

$$p\left(y_{lable}\middle|y_{pred}\right) = \frac{1}{Z(x)}\exp\left(\sum_{i,k}\lambda_k t_k\left(y_{label}^{i-1}, y_{label}^i, y_{pred}, i\right) + \sum_{i,l}\mu_l s_l\left(y_{label}^i, y_{pred}, i\right)\right) \tag{15}$$

where $Z\left(y_{pred}\right) = \sum_y \exp\left\{\sum_{i=1}^n w_i f_i\left(y_{lable}, y_{pred}\right)\right\}$ denotes the normalized factor, t_k is the transfer characteristic function, s_l is the transfer characteristic function, and λ_k and μ_l are the corresponding weight coefficients. In this study, the maximum likelihood estimation method was used to optimize the CRF. The likelihood function is given as follows:

$$L(W) = \sum_i \log p\left(y_{label}^i\middle|y_{label}^i\right) \tag{16}$$

The goal of optimization is to obtain the maximum conditional probability of the above formula. This study used the Viterbi algorithm to optimize the CRF calculation process. The final prediction result y_{hidden} is obtained from the output y_{pred} of the fully connected layer and the conditional probability $p\left(y_{lable}\middle|y_{pred}\right)$ calculated using the CRF.

2.5.4. Loss Function

A cross entropy is introduced as the loss function of the entire model to achieve optimization and convergence. The loss function based on cross entropy is written as Equation (17). The reasons for choosing cross entropy in this study are listed as follows: (1) Because the difference in the original data of the landslide factor is relatively small, the logarithmic function in the cross entropy is used to expand the gap between the data, thereby reducing the numerical calculation error. (2) Because the influences of different influence factors on the landslide are different, the cross entropy is introduced to ensure that each impact factor has a different weight. The model can find the direction of the fastest gradient descent and accelerate the convergence of the model. (3) Cross entropy makes the probability of optimization more accurate, i.e., the probability of a landslide is closer to 1, and the probability of a non-landslide is closer to zero.

$$L_{\text{cross-entropy}} = -\frac{1}{n}\sum_{i=1}^n \left(y_{label}^i \log(y_{hidden}^i) + (1 - y_{label}^i)\log(1 - y_{hidden}^i)\right) \tag{17}$$

3. Results

3.1. Landslide-Related Spatial Dataset

3.1.1. Landslide-Related Environmental Factors

In the existing study, no agreement was found regarding the causes of landslide events due to the complexity of the geological environment and the sensitivity of landslide evolution [69]. However, based on studies between landslide and environmental factors conducted by researchers in different regions, the environmental factors consist of five main categories: topography and geomorphology, hydrologic environment, basic geology, land cover, and human activities [70]. Therefore, 14 environmental factors were selected as input variables for landslide prediction in this study: elevation, slope, aspect, plan curvature, profile curvature, relief amplitude, *NDBI*, total surface radiation, *NDVI*, population density index, distances to rivers, *MNDWI*, *TWI*, and lithology (Figure 4).
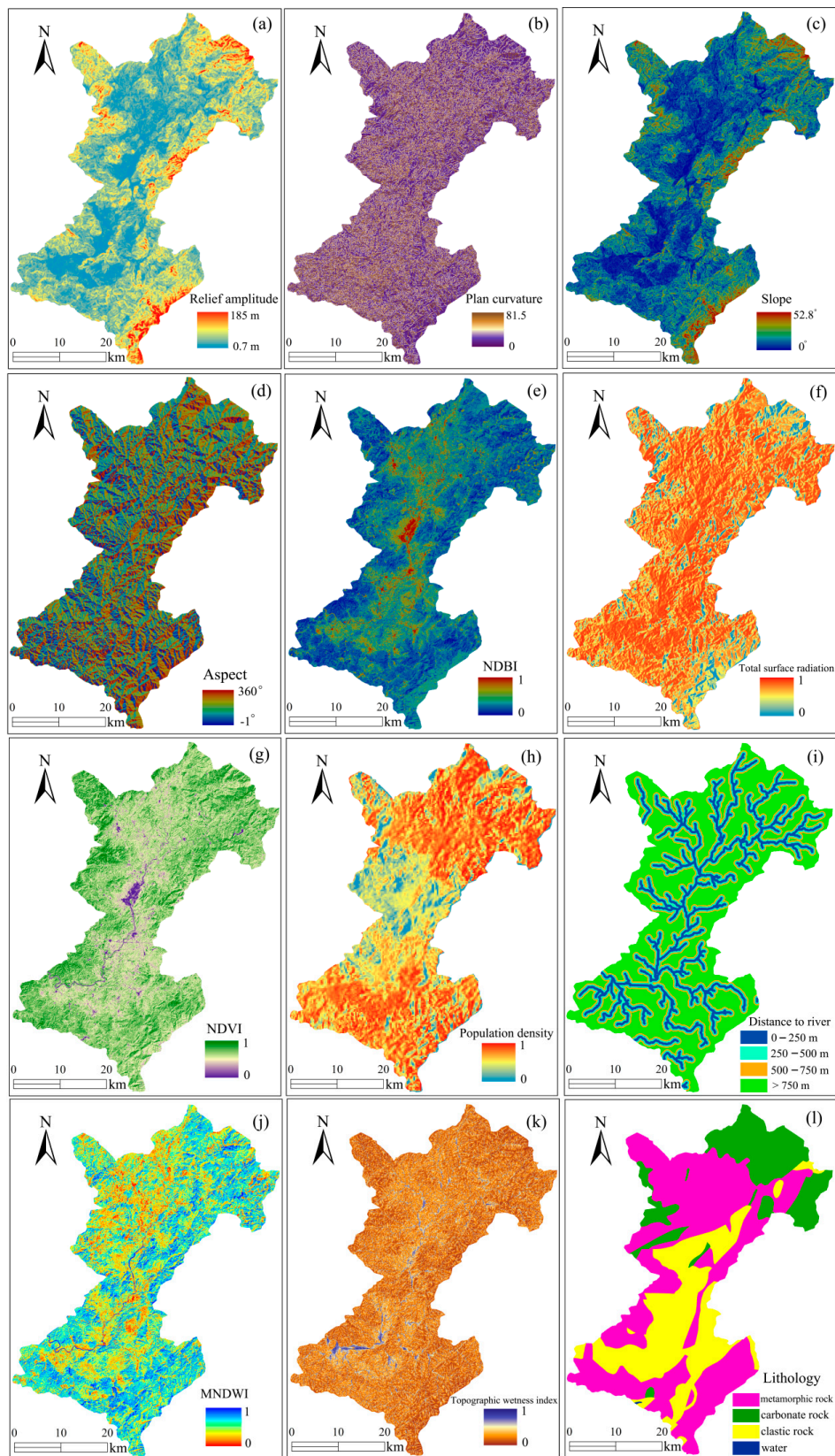
Figure 4. Landslide environmental factors (elevation and profile curvature are not presented): (**a**) Relief amplitude, (**b**) Plan curvature, (**c**) Slope, (**d**) Aspect, (**e**) NDBI, (**f**) Total surface radiation, (**g**) NDVI, (**h**) Population density, (**i**) Distance to river, (**j**) MNDWI, (**k**) Topographic wetness index, (**l**) Lithology.

It is significant to determine an appropriate spatial resolution of grid cells for predicting landslide susceptibility. This is because too low of a spatial resolution cannot ensure the reliability of the LSP while too high of a spatial resolution will strongly increase the complexity of the LSP modeling [71,72]. In this study, the original spatial resolutions of the grid cells of the DEM and Landsat TM images were both 30 m. This resolution value can effectively characterize the topography of Shicheng County and can avoid excessive computation [73]. In addition, a lot of literature suggests that a spatial resolution of 30 m is feasible and satisfactory for LSP [6,30,48,70,74,75]. Hence, the spatial resolution of grid cells in this study was set to 30 m.

3.1.2. Frequency Ratio Values of Environmental Factors

The frequency ratio values of these environmental factors were calculated to express the nonlinear correlations between environmental factors and landslides, as shown in Figure 5. The relief amplitude often appears in the range of 22–71. The plan curvature reflects the convergence and divergence of surface water flow [34], and the region of 0–27.5 is indicative of landslide events. The slope shows the difference in the surface water collecting capacity [76]. In addition, the aspect is closely related to landslide development and is often considered in the LSP, which is prone to landslide events in the ranges of 67.5–157.5 and 247.5–292.5. *NDBI* and population density intensively embody the range of human engineering activities with a high landslide probability. *NDVI* and total surface radiation reflect the impact of land cover on landslides [1], and an *NDVI* of 0.2–0.59 and a total surface radiation of 0.46–0.59 indicate a favorable environment for landslide occurrence. Moreover, the distances to rivers are the most common hydrologic environmental factor that affects the development of landslides, and the results show that the area with a distance to rivers within 250 m had the densest landslides. *MNDWI* was also selected as an environmental factor, and the landslides were mostly distributed in the range of 0.145–0.612. *TWI* reflects the centralized distribution state of water on the surface [77], and many landslides occur in range of 6.1–9.6. In addition, landslides are more likely to occur in the area with clastic rock than areas with metamorphic and carbonate rocks.

3.1.3. Model Training and Testing Datasets

It is necessary to prepare the training dataset and the test dataset for model training and testing prior to the prediction and modeling of landslide susceptibility. In this study, 369 landslides were converted into 2709 grid points in ARCGIS 10.2 software and randomly divided such that 70% and 30% went toward constructing the training and test sets, respectively. The same number of non-landslides were randomly divided using the same proportions to satisfy the binary classification condition. In addition, the *FR* values of the 14 environmental factors were used as input variables for the machine learning models, and the corresponding landslide and non-landslide results were marked as 1 and 0, which were treated as model output variables when constructing the dataset.
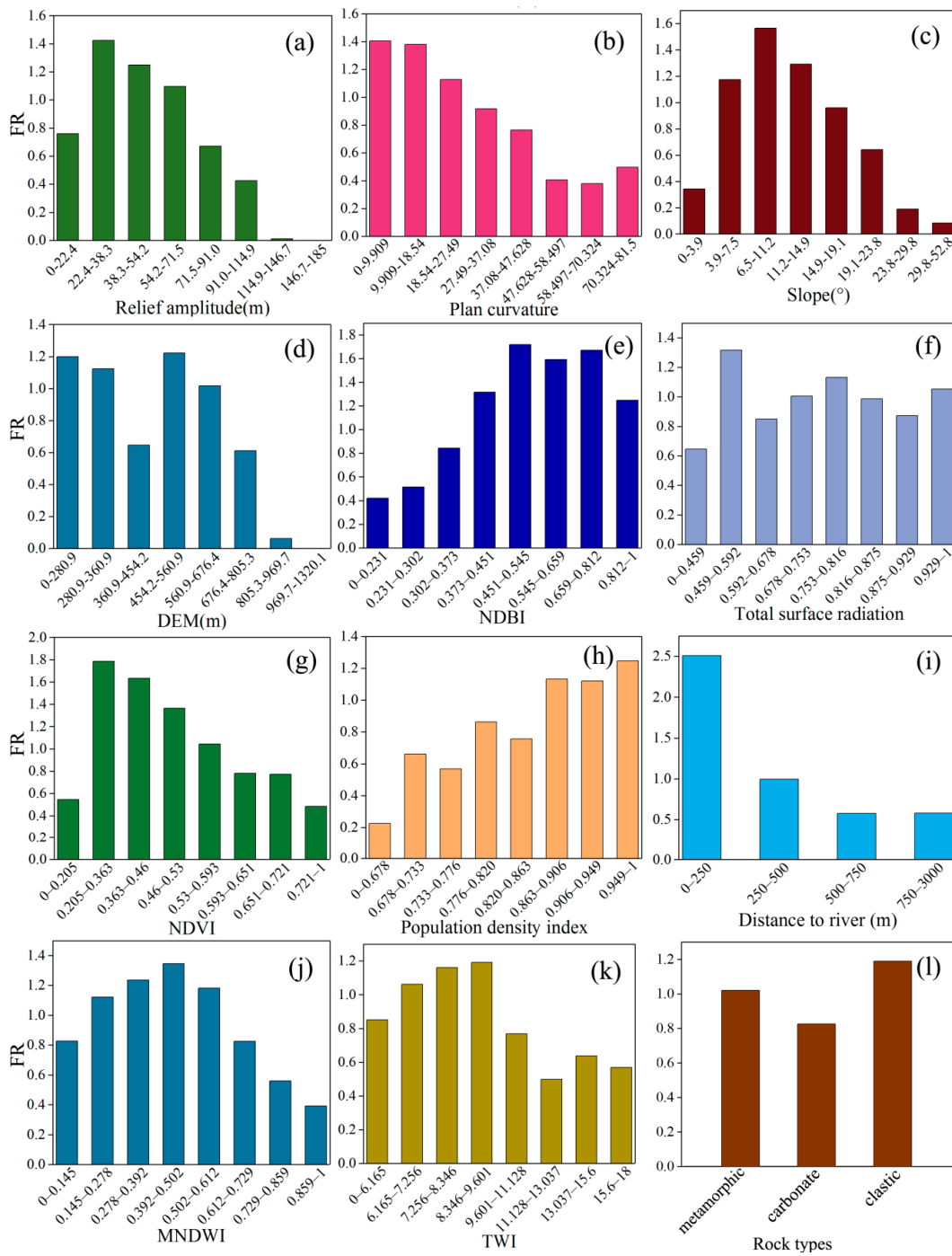
Figure 5. Frequency ratio values of environmental factors (aspect and profile curvature are not presented).

3.2. LSP Results of Shicheng County

The LSP of Shicheng County was implemented using each of the cascade-parallel LSTM-CRF, LR, C5.0 DT, and MLP models.

3.2.1. LSP Using the Cascade-Parallel LSTM-CRF Model

The hardware configuration required for the cascade-parallel LSTM-CRF model in the operating environment is shown in Table 1. In this study, the hyper-parameter search was adopted to acquire the optimal parameters in the cascade-parallel LSTM-CRF, where the number of LSTM units, learning rate, and total training parameters were 5, 0.001, and 40,432, respectively. The trained cascade-parallel

LSTM-CRF was used to predict the landslide susceptibility indices (LSIs) (Figure 7a) and to obtain the LSM (Figure 6a). In addition, the study area was divided into five categories, namely extremely low, low, moderate, high, and extremely high, using Jenks natural breaks optimization, and the corresponding area proportions were 33.56%, 25.41%, 18.33%, 10.98%, and 11.72%, respectively (Figure 7a).

Table 1. Software environment and hardware configuration in the experiment.

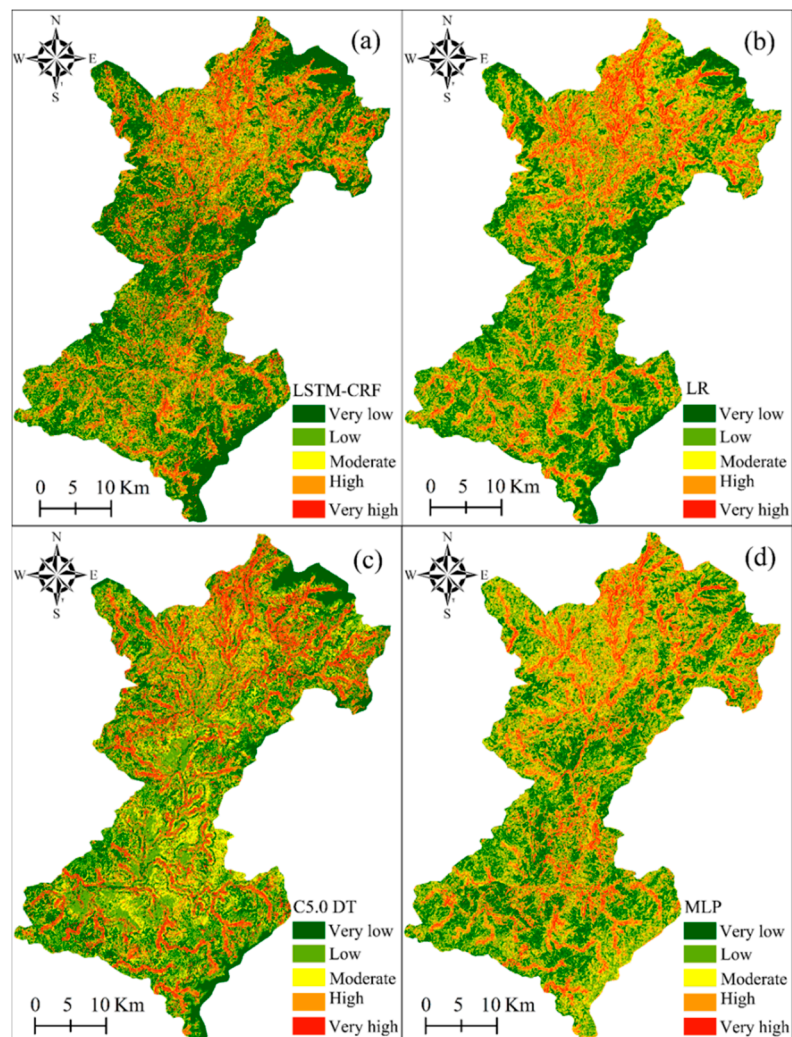| Software/Hardware	Parameters
Software Environment	TensorFlow1.4.0 + Keras2.1.4 (for cascade-parallel LSTM-CRF)
	SPSS Modeler 18.0 + Matlab 2018 (for logistic regression, multilayer perceptron, and C5.0 decision tree)
CPU	Intel(R) Core(TM) i5-7400@3.00 GHz
GPU	Nvidia GeForce GTX1080
RAM	8.00 GB DDR3
Hard Disk	Western Digital WDC WD10EZEX-08WN4A0

Figure 6. Landslide susceptibility map (LSM) produced by the four models. LR: Logistic regression, C5.0 DT: C5.0 decision tree, MLP: Multilayer perceptron.
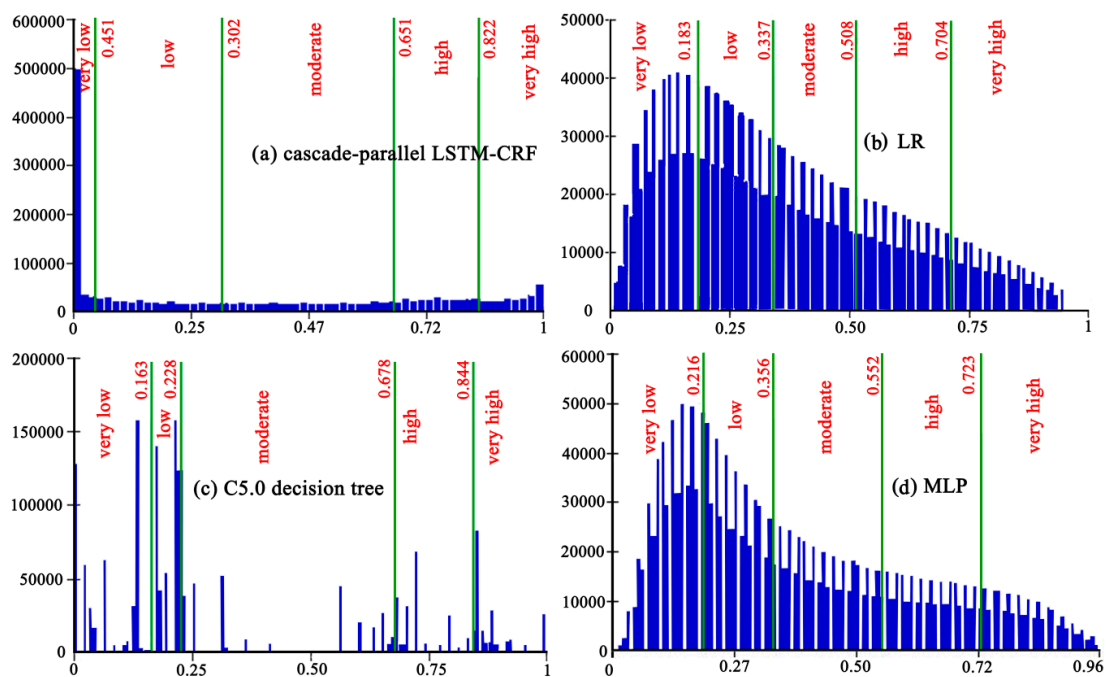
Figure 7. Landslide susceptibility index (LSI) distribution features of (**a**) LR, (**b**) MLP, (**c**) C5.0 decision tree, and (**d**) cascade-parallel LSTM-CRF.

3.2.2. LSP Using the LR Model

LR is a common statistical model that is widely used in LSP. The advantage of LR is that the independent variables need not be normally distributed, and the dependent variables can be continuous, discrete, and binary [78]. LR can define the relationship between the landslide occurrence and environmental factors and give the landslide coefficient of the corresponding factors. The logistic regression model can be expressed as follows:

$$z = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{18}$$

$$P = ln(\frac{p}{1-p}) = \frac{1}{1+e^{-z}}, \tag{19}$$

where z is the dependent variable, including landslides (1) and non-landslides (0); a_0 is the regression intercept; $x_n(n = 1, 2, \cdots, n)$ is the environmental factor; $a_n(n = 1, 2, \cdots, n)$ is the regression coefficient; and P is the probability of landslide occurrence with a range from 0 to 1.

Considering the information redundancy between factors, 14 input variables were adopted to conduct correlation analysis in SPSS 23.0 software, and the results showed that the correlation between the factors was weak. All factors were imported into the LR model for calculation, and the profile curvature, relief amplitude, and total surface radiation were removed in the first calculation because the significance value was greater than 0.05. The remaining 11 environmental factors were imported into LR again, and the results are shown in Table 2. The regression coefficients of all factors were positive, indicating that each factor had a catalytic effect on the generation of landslides. Furthermore, lithology was the most influential factor for landslides in Shicheng County, followed by slope, *MNDWI*, plane curvature, population density, elevation, and *NDBI*. The regression coefficients of each factor were used to obtain the LSIs (Figure 7b) of the study area to generate the LSM (Figure 6b). The research area was divided into the five categories of extremely low, low, moderate, high, and extremely high, with the corresponding area proportions of 28.64%, 23.81%, 21.33%, 13.43%, and 12.79%, respectively (Figure 7b).

Table 2. The parameters in the LR model.

Environmental Factors	B	S.E.	Wald	Sig.	Exp (B)
Elevation	0.827	0.145	32.406	0	2.287
Slope	1.276	0.080	252.078	0	3.583
Aspect	0.691	0.159	18.917	0	1.996
Plan curvature	1.088	0.093	138.101	0	2.968
MNDWI	1.093	0.121	81.136	0	2.983
Distance to river	0.627	0.039	264.112	0	1.872
TWI	0.624	0.177	12.462	0	1.866
Lithology	2.397	0.269	79.468	0	10.990
NDVI	0.402	0.141	8.156	0.004	1.494
NDBI	0.810	0.102	62.701	0	2.249
Population density	0.899	0.150	36.063	0	2.457
Constant	−11.385	0.497	524.967	-	-

B is the regression coefficient, S.E. is the standard error, Wald is the Wald chi-square test, Sig. is the significance of the regression coefficient, and Exp(B) is the power of the regression coefficient.

3.2.3. LSP Using the C5.0 DT Model

The C5.0 DT is a tree structure that can intuitively interpret decision rules based on input conditions. The sample data were segmented using the input variable with the highest information gain in entropy, and the information gain reflects the change of variable information before and after the data are divided. The smaller the new conditional entropy change and the greater the information gain, the better the prediction accuracy of the model [79,80]. Moreover, the pruning of each leaf node is crucial to improving the model accuracy, which is beneficial for removing the environmental factors that have an insignificant effect on the decision results. The lower the purity of each leaf, the higher the classification accuracy. Additionally, the tree growth is limited with the pruning degree to avoid over-fitting of the model and to obtain a concise and accurate model.

In this study, the modeling of C5.0 DT was conducted in SPSS Modeler 18.0 software, the main contents of which included the input of variables, tree growth, tree pruning, and model validation. The boosting algorithm was adopted to improve the model prediction accuracy, and the pruning degree and the number of each branch node are the default values. The LSIs were calculated using the trained C5.0 DT, as shown in Figure 7c. The LSM generated in Shicheng County is shown in (Figure 6c). Furthermore, the study area was divided into extremely low, low, moderate, high, and extremely high areas, with the corresponding areas accounting for 30.38%, 24.13%, 19.74%, 12.19%, and 13.55%, respectively (Figure 7c).

3.2.4. LSP Using the MLP Model

MLP is an artificial neural network technology that has been widely used in pattern recognition and classification [81], and it is often adopted for solving the complicated nonlinear problem between landslides and various environmental factors in the process of LSP. In addition, the MLP's advantage is its powerful ability to work with inaccurate and fuzzy data [82], and the main components of MLP include an input layer, a hidden layer containing at least one neuron, and an output layer. Random initial weights are assigned among all neuron nodes, and the hidden layer does not contact the external environment, which addresses the sum of the product of input values and connection weights between each layer with a nonlinear function.

The neural network algorithm simulates human learning primarily by strengthening and adjusting the connection weight between the input layer and the output layer, and the performance in the network is used to update the internal system structure due to an external stimulus [83]. In this study, an extensively applied backpropagation algorithm was adopted to construct the MLP neural network. MLP neural network modeling consists of three stages: (1) in the process of the forward transmission of the input data, the sigmoid function is used to process the sum of the products of the weights between layers, and the error between the output value and the theoretical value is calculated; (2) in the iterative process of backward transmission, the calculated error value is used to continuously update the connection weights to obtain the optimal error value; and (3) the trained network structure is used in the classification and prediction of continuous data in the entire research area.

The number of required neurons in a single hidden layer was calculated to be 25 based on the minimum prediction error method. Therefore, the numbers of neurons in the input layer, the hidden layer, and the output layer were 14, 25, and 1, respectively, and via different tests, the momentum, learning rate, and training time were 0.36, 0.05, and 500, respectively. Moreover, the trained MLP model was used to calculate the LSIs (Figure 7d) and to acquire the LSM (Figure 6d), which was divided into the five categories of extremely low, low, moderate, high, and extremely high, with the corresponding area proportions of 27.7%, 22.9%, 22.7%, 16.9%, and 9.9%, respectively (Figure 7d).

3.3. Accuracy Evaluation

Quality evaluation of the models is a key step in successful landslide prediction. In this study, the predictive rate curves and statistical indicators were used to assess the predictive performance of the models.

3.3.1. Predictive Rate Accuracy

The performance evaluation of the model is related to the model's application of LSP in the study area [84]. To analyze and compare the prediction ability of each model, the prediction rate curve was adopted to evaluate the fitting degree between landslide grid cells in the testing dataset and the predicted landslide susceptibility indices in the study area. First of all, the calculated LSI values were sorted in descending order. Second, these sorted LSI values were categorized into 20 equal intervals with 5% cumulative intervals. Third, the percentage of the landslide grid cells in the testing dataset of each interval was determined based on the former obtained 20 equal intervals. At last, the prediction rate curves of all four models were drawn. The area under the predictive rate curve (AUC) could clearly explain the LSP performance of the model with the threshold value, which was closer to 1, indicating that the model's performance was better.

The predictive rate curves for different models are shown in Figure 8. In the top 10% of the very high susceptibility index interval, the cumulative percentages of landslide grids in the test set of the cascade-parallel LSTM-CRF, C5.0 DT, LR, and MLP models were 44.04%, 29.84%, 43.85%, and 43.81%, respectively. Moreover, in the top 20% of the study area with a high susceptibility index and above, the percentages of landslide grids in the corresponding models were 67.48%, 63.58%, 61.68%, and 61.43%, respectively. The prediction rate curve intuitively reflected the fact that the cascade-parallel LSTM-CRF, LR, and MLP models had a positive prediction performance in very highly prone areas. Overall, the AUC values of the cascade-parallel LSTM-CRF, C5.0 DT, LR, and MLP models were 0.868, 0.838, 0.833, and 0.826, respectively, indicating that the cascade-parallel LSTM-CRF model had a better predictive ability than other models.
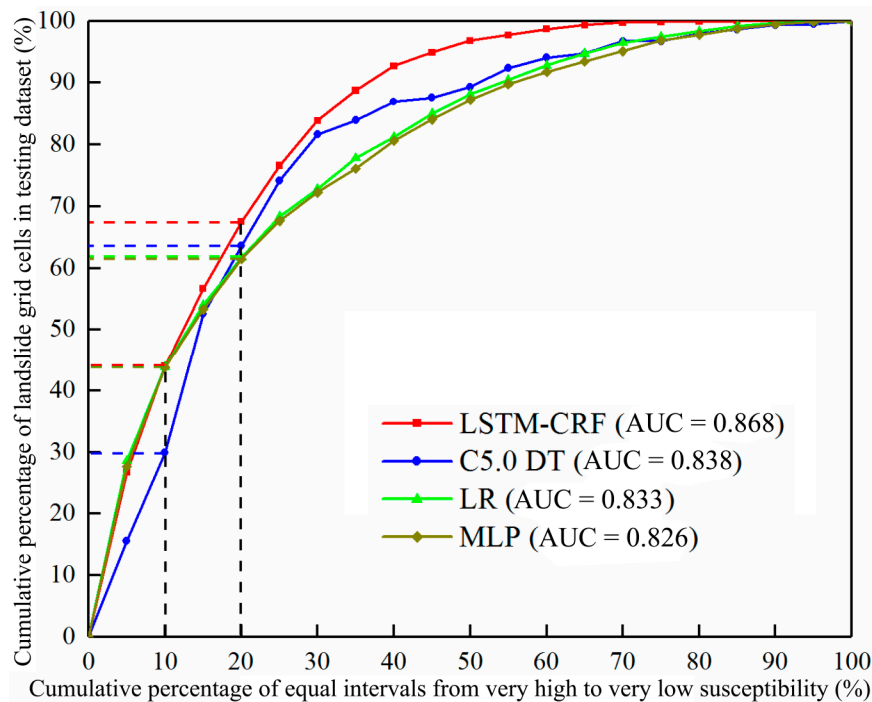
Figure 8. Prediction rate curves of LSIs calculated by the four models.

3.3.2. Statistical Index Accuracy

In this study, the statistical indexes, including the positive predictive rate (*PPR*), negative predictive rate (*NPR*), and total accuracy (*TA*), were used to measure the predictive performance of the models. The positive and negative predictive rates respectively represent the proportion of landslides and non-landslides that were correctly classified in the actual statistical process. Total accuracy represents the total predictive rate with which landslides and non-landslides are correctly classified. The corresponding expressions are given as:

$$PPR = \frac{TP}{TP + FP} \tag{20}$$

$$NPR = \frac{TN}{TN + FN} \tag{21}$$

$$TA = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

where *TP* and *FP* respectively represent the number of landslide and non-landslide grids that are classified correctly, and *TN* and *FN* respectively represent the number of landslide and non-landslide grids that are classified mistakenly.

The performance of each model obtained using statistical indicators is presented in Table 3. In the predictive statistics process, the positive prediction rate of the C5.0 DT model produced the highest proportion of correct landslide classifications (76.78%), followed by the cascade-parallel LSTM-CRF model (72.44%), the MLP model (71.58%), and the LR model (71.04%). Regarding correct non-landslide classification, the cascade-parallel LSTM-CRF model produced the highest negative predictive rate of 80%, and those of the MLP, LR, and C5.0 DT models were 71.46%, 70.83%, and 69.73%, respectively. The cascade-parallel LSTM-CRF model has the best prediction performance regarding both landslides and non-landslides, with a total overall accuracy of 75.69%, followed by C5.0 DT (72.72%), MLP (71.61%), and LR (70.94%). In terms of the classification accuracy of landslides and non-landslides, the cascade-parallel LSTM-CRF model was more suitable than the other traditional models for LSPs.

Table 3. Comparison between Landslide prediction performances.

| Prediction Performance	Prediction Models			
	Cascade-Parallel LSTM-CRF	C5.0 DT	LR	MLP
True positive	673	529	574	582
True negative	556	652	578	581
False positive	256	160	234	231
False negative	139	283	238	230
PPR (%)	72.44	76.78	71.04	71.58
NPR (%)	80.00	69.73	70.83	71.64
TR (%)	75.67	72.72	70.94	71.61

4. Discussion

4.1. Discussion of the LSP Model Accuracy

The prediction rate and statistical indexes of the model accuracies evaluation show that the cascade-parallel LSTM-CRF model had the highest LSP performance, followed by the C5.0 DT, LR, and MLP models. However, it is not guaranteed that the deep learning model in this paper, among all machine learning models, is the optimal choice; however, the comparison of the models can reflect the ability of the proposed model to indeed overcome the shortcomings of traditional machine learning models. The cascade-parallel LSTM is one of the better performing models among all machine learning models. In further research, we aim to compare as many other machine learning models as possible to find the shortcomings of the proposed model and improve the model as much as possible to improve the accuracy of the landslide susceptibility predictions.

4.2. Model Iteration and Accuracy

The loss is defined as the error between the theoretical value and the predicted value, which decreases with the updating of weights during backpropagation. The accuracy refers to the degree to which the predicted value is close to the theoretical value. The variation curves of the loss and accuracy in the model with a varying number of iterations are shown in Figure 9. The loss dropped sharply within 4000 iterations and stabilized at approximately 0.3 after 9000 iterations. During the same period, the accuracy of the model increased from 0.7 to 0.85, increased slowly until it became stable, and the final accuracy reached approximately 0.868.
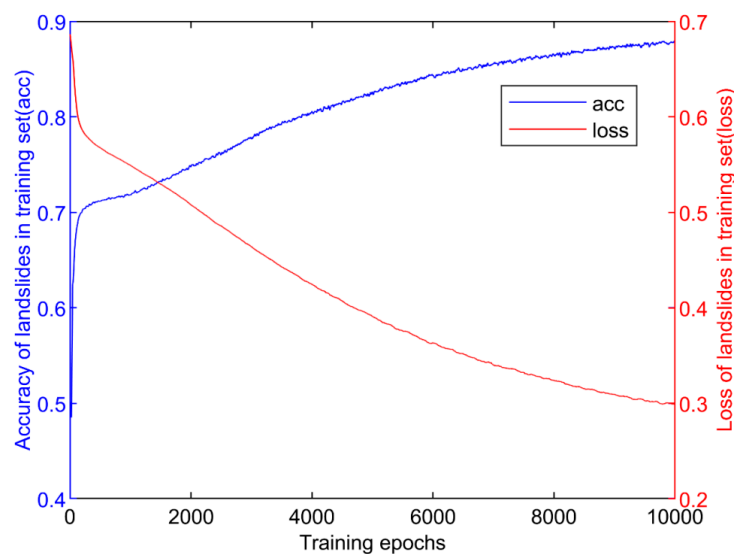
Figure 9. Loss and accuracy curves throughout the training of the models.

4.3. Distribution Characteristics of Landslide Susceptibility

Landslide susceptibility maps obtained from the above four models are shown in Figure. According to the susceptibility index of the entire study area, the distribution of the susceptibility calculated by each model was roughly similar, and the distribution rules were consistent with the actual geological environment in the study area. According to superposition analysis of the landslide susceptibility maps and the environmental factor maps, it was observed that very high and high prone areas were mainly distributed near rivers, where *NDBI* and population density were relatively high, indicating that human engineering activities had a great impact on landslides. Moreover, most of the landslides were distributed in areas with a small *NDVI* and a large total surface radiation, which indicates that vegetation coverage had a certain inhibitory effect on landslide occurrence. Furthermore, landslides were always concentrated in brittle rock areas, such as areas with metamorphic and clastic rocks, which can create a favorable environment for the development of landslide events.

In contrast, very low and low-landslide-prone areas were primarily distributed in areas far from rivers and carbonate rocks, which had relatively dense vegetation coverage, poor impounding conditions, less human activity, and a low probability of landslide occurrence. In addition, carbonate rocks were more conducive to slope stability and not conducive to landslides compared with metamorphic rocks and clastic rocks.

4.4. Advantages of the Cascade-Parallel LSTM-CRF Model

A cascade-parallel LSTM-CRF model for landslide prediction based on deep learning was proposed in this study. Compared with other traditional prediction models, this model has the following advantages in LSP: (1) The prediction accuracy obtained by this algorithm was high, which was mainly embodied in the correct classification stage of landslides and non-landslides. The total predictive rate in the process of training was 75.67%, and the precision of statistical indicators in the test set was 0.868, both higher than those of the C5.0 DT, LR, and MLP models. (2) This algorithm is mainly based on deep learning and is data-driven, which overcomes the limitations of traditional prediction models related to input data, such as the need for substantial prior knowledge, the satisfaction of certain distribution characteristics, mutual independence between factors, etc. Hence, the algorithm greatly improves the applicability in landslide prediction and has a high practical value. (3) The main feature of this algorithm is the adoption of the cascade-parallel LSTM structure, which uses the perception range characteristics of different LSTM structures as data features and extracts and fuses the features of different input factors to acquire deeper and more complete features of each factor; furthermore, dropout is adopted to prevent over-fitting in the training process. (4) Considering that the susceptibility of a grid cell is affected by the susceptibility results of adjacent grids, this study introduced CRF, and the cross entropy function is used to change the direction of the gradient descent in the algorithm and accelerate the model convergence to optimize the model.

5. Conclusions

In this study, the cascade-parallel LSTM-CRF model was constructed by introducing a deep learning algorithm and combining it with CRF to predict landslide susceptibility in Shicheng County based on remote sensing images and GIS. The cascade-parallel LSTM-CRF model was compared with traditional prediction models, and 14 environmental factors were chosen based on the geological conditions relevant to landslide occurrence. The models were trained and tested based on adopting data samples, which were randomly split using a 70/30 training/testing ratio. Finally, the predictive rate curve and statistical indices were used to evaluate the performance of each model.

The results show that the cascade-parallel LSTM-CRF model had a better prediction performance than other models. The advantages of this algorithm are listed as follows: (1) The proposed algorithm has a higher LSP performance than traditional machine learning algorithms. (2) The proposed algorithm overcomes the limitations associated with traditional machine learning algorithms requiring substantial

prior knowledge. Therefore, this algorithm is universal and practical, and can be applied to a wider range of landslide data predictions. (3) The stacked structure was adopted in this algorithm. Using the cascade-parallel LSTM, different layers of features can be extracted and merged from different layers, such as concrete and abstract. At the same time, the strategy of preventing over-fitting of the network was adopted, such that it can extract more comprehensive and accurate landslide features that facilitate classification. (4) CRF was further introduced in the cascade-parallel LSTM in this study. The predicted results are optimized by calculating the energy relationship between the two grid points, optimizing the extracted features and smoothing the predicted results of mutations.

Moreover, the distribution results of landslide susceptibility of each model were consistent with the actual geological environment in the study area. In conclusion, the introduction of the deep learning algorithm is of great significance to the prediction of landslide susceptibility, which can supply the local government with theoretical guidance for the rational allocation of land resources and the implementation of disaster prevention and mitigation measures.

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

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

References

1. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [CrossRef]

2. Cullen, C.A.; Al-Suhili, R.; Khanbilvardi, R. Guidance index for shallow landslide hazard analysis. *Remote Sens.* **2016**, *8*, 866. [CrossRef]

3. Wang, L.-J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. *Catena* **2015**, *135*, 271–282. [CrossRef]

4. Huang, F.; Yin, K.; Huang, J.; Gui, L.; Wang, P. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Eng. Geol.* **2017**, *223*, 11–22. [CrossRef]

5. Chang, Z.; Du, Z.; Zhang, F.; Huang, F.; Chen, J.; Li, W.; 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]

6. Shahabi, H.; Khezri, S.; Ahmad, B.B.; Hashim, M. Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena* **2014**, *115*, 55–70. [CrossRef]

7. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. *Geomorphology* **2012**, *171–172*, 12–29. [CrossRef]

8. Moayedi, H.; Osouli, A.; Tien Bui, D.; Foong, L.K. Spatial Landslide Susceptibility Assessment Based on Novel Neural-Metaheuristic Geographic Information System Based Ensembles. *Sensors* **2019**, *19*, 4698. [CrossRef]

9. Amatya, P.; Kirschbaum, D.; Stanley, T. Use of Very High-Resolution Optical Data for Landslide Mapping and Susceptibility Analysis along the Karnali Highway, Nepal. *Remote Sens.* **2019**, *11*, 2284. [CrossRef]

10. Zhao, C.; Lu, Z. Remote Sensing of Landslides—A Review. *Remote Sens.* **2018**, *10*, 279. [CrossRef]

11. Zhang, B.; Zhang, L.; Yang, H.; Zhang, Z.; Tao, J. Subsidence prediction and susceptibility zonation for collapse above goaf with thick alluvial cover: A case study of the Yongcheng coalfield, Henan Province, China. *Bull. Eng. Geol. Environ.* **2016**, *75*, 1117–1132. [CrossRef]

12. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2015**, *13*, 839–856. [CrossRef]

13. Wang, Q.; Wang, Y.; Niu, R.; Peng, L. Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility Mapping in the Three Gorges Area, China. *Remote Sens.* **2017**, *9*, 938. [CrossRef]

14. Erener, A.; Mutlu, A.; Sebnem Düzgün, H. A comparative study for landslide susceptibility mapping using GIS-based multi-criteria decision analysis (MCDA), logistic regression (LR) and association rule mining (ARM). *Eng. Geol.* **2016**, *203*, 45–55. [CrossRef]

15. Sevgen, E.; Kocaman, S.; Nefeslioglu, H.A.; Gokceoglu, C. A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ann and random forest. *Sensors* **2019**, *19*, 3940. [CrossRef]

16. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.B.; Panahi, M.; Hong, H.; et al. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sens.* **2018**, *10*, 1527. [CrossRef]

17. 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* **2012**, *65*, 135–165. [CrossRef]

18. 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.* **2015**, *75*, 63. [CrossRef]

19. Hasekioğulları, G.D.; Ercanoglu, M. A new approach to use AHP in landslide susceptibility mapping: A case study at Yenice (Karabuk, NW Turkey). *Nat. Hazards* **2012**, *63*, 1157–1179. [CrossRef]

20. Nguyen, T.T.N.; Liu, C.-C. A New Approach Using AHP to Generate Landslide Susceptibility Maps in the Chen-Yu-Lan Watershed, Taiwan. *Sensors* **2019**, *19*, 505. [CrossRef]

21. Wang, Y.; Wu, X.; Chen, Z.; Ren, F.; Feng, L.; Du, Q. Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using smote for lishui city in zhejiang province, china. *Int. J. Environ. Res. Public Health* **2019**, *16*, 368. [CrossRef] [PubMed]

22. Yu, X.; Wang, Y.; Niu, R.; Hu, Y. A combination of geographically weighted regression, particle swarm optimization and support vector machine for landslide susceptibility mapping: A case study at Wanzhou in the Three Gorges Area, China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 487. [CrossRef] [PubMed]

23. Chen, W.; Hong, H.; Panahi, M.; Shahabi, H.; Wang, Y.; Shirzadi, A.; Pirasteh, S.; Alesheikh, A.A.; Khosravi, K.; Panahi, S. Spatial prediction of landslide susceptibility using gis-based data mining techniques of anfis with whale optimization algorithm (woa) and grey wolf optimizer (gwo). *Appl. Sci.* **2019**, *9*, 3755. [CrossRef]

24. Truong, X.L.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, X.Q.; Do, T.H.; Tien Bui, D.; Lee, S. Enhancing Prediction Performance of Landslide Susceptibility Model Using Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree. *Appl. Sci.* **2018**, *8*, 1046. [CrossRef]

25. Conoscenti, C.; Ciaccio, M.; Caraballo-Arias, N.A.; Gómez-Gutiérrez, Á.; Rotigliano, E.; Agnesi, V. Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Belice River basin (western Sicily, Italy). *Geomorphology* **2015**, *242*, 49–64. [CrossRef]

26. Felicísimo, Á.M.; Cuartero, A.; Remondo, J.; Quirós, E. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslides* **2012**, *10*, 175–189. [CrossRef]

27. Zhu, A.X.; Wang, R.; Qiao, J.; Qin, C.-Z.; Chen, Y.; Liu, J.; Du, F.; Lin, Y.; Zhu, T. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic. *Geomorphology* **2014**, *214*, 128–138. [CrossRef]

28. Huang, F.; Huang, J.; Jiang, S.; Zhou, C. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine. *Eng. Geol.* **2017**, *218*, 173–186. [CrossRef]

29. Bui, D.T.; Moayedi, H.; Kalantar, B.; Osouli, A.; Pradhan, B.; Nguyen, H.; Rashid, A.S.A. A novel swarm intelligence—harris hawks optimization for spatial assessment of landslide susceptibility. *Sensors* **2019**, *19*, 3590. [CrossRef]

30. Li, D.; Huang, F.; Yan, L.; Cao, Z.; Chen, J.; Ye, Z. Landslide Susceptibility Prediction Using Particle-Swarm-Optimized Multilayer Perceptron: Comparisons with Multilayer-Perceptron-Only, BP Neural Network, and Information Value Models. *Appl. Sci.* **2019**, *9*, 3664. [CrossRef]

31. Tsangaratos, P.; Ilia, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* **2015**, *13*, 305–320. [CrossRef]

32. Shirzadi, A.; Soliamani, K.; Habibnejhad, M.; Kavian, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [CrossRef] [PubMed]

33. Park, S.-J.; Lee, C.-W.; Lee, S.; Lee, M.-J. Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea. *Remote Sens.* **2018**, *10*, 1545. [CrossRef]

34. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [CrossRef]

35. Lai, J.-S.; Tsai, F. Improving GIS-based landslide susceptibility assessments with multi-temporal remote sensing and machine learning. *Sensors* **2019**, *19*, 3717. [CrossRef]

36. Park, S.; Kim, J. Landslide Susceptibility Mapping Based on Random Forest and Boosted Regression Tree Models, and a Comparison of Their Performance. *Appl. Sci.* **2019**, *9*, 942. [CrossRef]

37. Dai, S.; Niu, D.; Han, Y. Forecasting of power grid investment in china based on support vector machine optimized by differential evolution algorithm and grey wolf optimization algorithm. *Appl. Sci.* **2018**, *8*, 636. [CrossRef]

38. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A novel ensemble approach for landslide susceptibility mapping (lsm) in darjeeling and kalimpong districts, west bengal, india. *Remote Sens.* **2019**, *11*, 2866. [CrossRef]

39. Huang, F.; Yin, K.; Zhang, G.; Zhou, C.; Zhang, J. Landslide groundwater level time series prediction based on phase space reconstruction and wavelet analysis-support vector machine optimized by pso algorithm. *Earth Sci.-J. China Univ. Geosci.* **2015**, *40*, 1254–1265.

40. Roodposhti, M.S.; Aryal, J.; Pradhan, B. A novel rule-based approach in mapping landslide susceptibility. *Sensors* **2019**, *19*, 2274. [CrossRef]

41. Nsengiyumva, J.B.; Luo, G.; Nahayo, L.; Huang, X.; Cai, P. Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation Model in Rwanda. *Int. J. Environ. Res. Public Health* **2018**, *15*, 243. [CrossRef] [PubMed]

42. Tien Bui, D.; Hoang, N.D.; Martinez-Alvarez, F.; Ngo, P.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2019**, *701*, 134413. [CrossRef] [PubMed]

43. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]

44. Liu, Y.; Wu, L. Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning. *Procedia Comput. Sci.* **2016**, *91*, 566–575. [CrossRef]

45. Hu, G.; Yang, Y.; Yi, D.; Kittler, J.; Christmas, W.; Li, S.Z.; Hospedales, T. When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 7–13 December 2015; pp. 384–392.

46. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131.e9. [CrossRef]

47. Xiao, L.; Zhang, Y.; Peng, G. Landslide Susceptibility Assessment Using Integrated Deep Learning Algorithm along the China-Nepal Highway. *Sensors* **2018**, *18*, 4436. [CrossRef]

48. Huang, F.; Zhang, J.; Zhou, C.; Wang, Y.; Huang, J.; Zhu, L. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* **2020**, *17*, 217–229. [CrossRef]

49. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

50. He, K.; Zhang, X.; Ren, S. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

51. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]

52. Hoek, E.; Brown, E.T. Practical estimates of rock mass strength. *Int. J. Rock Mech. Min. Sci.* **1997**, *34*, 1165–1186. [CrossRef]

53. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*; Springer: New York, NY, USA, 2002.

54. Cox, D.R. The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B (Methodol.)* **1958**, *20*, 215–232. [CrossRef]

55. Ripley, B.D. *Pattern Classification and Neural Networks*; Cambridge University Press: Cambridge, UK, 1996.

56. Hungr, O.; Leroueil, S.; Picarelli, L. The Varnes classification of landslide types, an update. *Landslides* **2014**, *11*, 167–194. [CrossRef]

57. Lin, G.F.; Chang, M.J.; Huang, Y.C.; Ho, J.Y. Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. *Eng. Geol.* **2017**, *224*, 62–74. [CrossRef]

58. Szymura, T.H.; Szymura, M.; Pietrzak, M. Influence of land relief and soil properties on stand structure of overgrown oak forests of coppice origin with Sorbus torminalis. *Dendrobiology* **2014**, *71*, 49–58. [CrossRef]

59. Dixon, N.; Brook, E. Impact of predicted climate change on landslide reactivation: Case study of Mam Tor, UK. *Landslides* **2007**, *4*, 137–147. [CrossRef]

60. Duc, D.M. Rainfall-triggered large landslides on 15 December 2005 in Van Canh District, Binh Dinh Province, Vietnam. *Landslides* **2013**, *10*, 219–230. [CrossRef]

61. Sun, J.; Cheng, G.; Li, W.; Sha, Y.; Yang, Y. On the variation of ndvi with the principal climatic elements in the tibetan plateau. *Remote Sens.* **2013**, *5*, 1894–1911. [CrossRef]

62. Jia, N.; Mitani, Y.; Xie, M.; Tong, J.; Yang, Z. GIS deterministic model-based 3D large-scale artificial slope stability analysis along a highway using a new slope unit division method. *Nat. Hazards* **2015**, *76*, 873–890. [CrossRef]

63. Erener, A.; Düzgün, H.S.B. Landslide susceptibility assessment: What are the effects of mapping unit and mapping method? *Environ. Earth Sci.* **2012**, *66*, 859–877. [CrossRef]

64. Arabameri, A.; Pradhan, B.; Rezaei, K.; Lee, C.-W. Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs. *Remote Sens.* **2019**, *11*, 999. [CrossRef]

65. Tong, Q.; Li, X.; Lin, K. Cascade LSTM Based Visual-Inertial Navigation for Magnetic Levitation Haptic Interaction. *IEEE Netw.* **2019**, *33*, 74–80. [CrossRef]

66. Zhu, X.; Sobhani, P.; Guo, H. Long short-term memory over recursive structures. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015; pp. 1604–1612.

67. Zeng, D.; Sun, C.; Lin, L. LSTM-CRF for drug-named entity recognition. *Entropy* **2017**, *19*, 283. [CrossRef]

68. Gridach, M. Character-level neural network for biomedical named entity recognition. *J. Biomed. Inf.* **2017**, *70*, 85–91. [CrossRef]

69. Huang, F.; Luo, X.; Liu, W. Stability analysis of hydrodynamic pressure landslides with different permeability coefficients affected by reservoir water level fluctuations and rainstorms. *Water* **2017**, *9*, 450. [CrossRef]

70. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [CrossRef]

71. Shirzadi, A.; Solaimani, K.; Roshan, M.H.; Kavian, A.; Chapi, K.; Shahabi, H.; Keesstra, S.; Ahmad, B.B.; Bui, D.T. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *Catena* **2019**, *178*, 172–188. [CrossRef]

72. Arnone, E.; Francipane, A.; Scarbaci, A.; Puglisi, C.; Noto, L.V. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environ. Model. Softw.* **2016**, *84*, 467–481. [CrossRef]

73. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [CrossRef]

74. Huang, F.; Wang, Y.; Dong, Z.; Wu, L.; Guo, Z.; Zhang, T. Regional Landslide Susceptibility Mapping Based on Grey Relational Degree Model. *Earth Sci.* **2019**, *44*, 664–676.

75. Huang, F.; Yao, C.; Liu, W.; Li, Y.; Liu, X. Landslide susceptibility assessment in the Nantian area of China: A comparison of frequency ratio model and support vector machine. *Geomat. Nat. Hazards Risk* **2018**, *9*, 919–938. [CrossRef]

76. Hong, H.; Ilia, I.; Tsangaratos, P.; Chen, W.; Xu, C. A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. *Geomorphology* **2017**, *290*, 1–16. [CrossRef]

77. Liu, W.; Luo, X.; Huang, F.; Fu, M. Uncertainty of the soil–water characteristic curve and its effects on slope seepage and stability analysis under conditions of rainfall using the markov chain monte carlo method. *Water* **2017**, *9*, 758. [CrossRef]

78. Cao, J.; Zhang, Z.; Wang, C.; Liu, J.; Zhang, L. Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *Catena* **2019**, *175*, 63–76. [CrossRef]

79. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [CrossRef]

80. Tseng, C.-J.; Lu, C.-J.; Chang, C.-C.; Chen, G.-D.; Cheewakriangkrai, C. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. *Artif. Intell. Med.* **2017**, *78*, 47–54. [CrossRef] [PubMed]

81. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [CrossRef]

82. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [CrossRef]

83. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [CrossRef]

84. Cantarino, I.; Carrion, M.A.; Goerlich, F.; Martinez Ibañez, V. A ROC analysis-based classification method for landslide susceptibility maps. *Landslides* **2018**, *16*, 265–282. [CrossRef]