

## Article

# An Efficient User-Friendly Integration Tool for Landslide Susceptibility Mapping Based on Support Vector Machines: SVM-LSM Toolbox

Wubiao Huang <sup>1,2</sup>, Mingtao Ding <sup>1,2,3,\*</sup>, Zhenhong Li <sup>1,2,3</sup>, Jianqi Zhuang <sup>1,3</sup>, Jing Yang <sup>1,2</sup>, Xinlong Li <sup>1,2</sup>, Ling'en Meng <sup>1,2</sup>, Hongyu Zhang <sup>1,2</sup> and Yue Dong <sup>1,2</sup>

<sup>1</sup> College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China; huangwubiao@chd.edu.cn (W.H.); zhenhong.li@chd.edu.cn (Z.L.); jqzhuang@chd.edu.cn (J.Z.); 2020126052@chd.edu.cn (J.Y.); xinlong.li@chd.edu.cn (X.L.); 2020226049@chd.edu.cn (L.M.); 2021226062@chd.edu.cn (H.Z.); 2020126048@chd.edu.cn (Y.D.)

<sup>2</sup> Big Data Center for Geosciences and Satellites, Chang'an University, Xi'an 710054, China

<sup>3</sup> Key Laboratory of Western China's Mineral Resource and Geological Engineering, Ministry of Education, Xi'an 710054, China

\* Correspondence: mingtaodding@chd.edu.cn

**Abstract:** Landslide susceptibility mapping (LSM) is an important element of landslide risk assessment, but the process often needs to span multiple platforms and the operation process is complex. This paper develops an efficient user-friendly toolbox including the whole process of LSM, known as the SVM-LSM toolbox. The toolbox realizes landslide susceptibility mapping based on a support vector machine (SVM), which can be integrated into the ArcGIS or ArcGIS Pro platform. The toolbox includes three sub-toolboxes, namely: (1) influence factor production, (2) factor selection and dataset production, and (3) model training and prediction. Influence factor production provides automatic calculation of DEM-related topographic factors, converts line vector data to continuous raster factors, and performs rainfall data processing. Factor selection uses the Pearson correlation coefficient (PCC) to calculate the correlations between factors, and the information gain ratio (IGR) to calculate the contributions of different factors to landslide occurrence. Dataset sample production includes the automatic generation of non-landslide data, data sample production and dataset split. The accuracy, precision, recall, F1 value, receiver operating characteristic (ROC) and area under curve (AUC) are used to evaluate the prediction ability of the model. In addition, two methods—single processing and multiprocessing—are used to generate LSM. The prediction efficiency of multiprocessing is much higher than that of the single process. In order to verify the performance and accuracy of the toolbox, Wuqi County, Yan'an City, Shaanxi Province was selected as the test area to generate LSM. The results show that the AUC value of the model is 0.8107. At the same time, the multiprocessing prediction tool improves the efficiency of the susceptibility prediction process by about 60%. The experimental results confirm the accuracy and practicability of the proposed toolbox in LSM.

**Keywords:** landslide susceptibility mapping; toolbox; SVM; automatic; multiprocessing; the whole process

**Citation:** Huang, W.; Ding, M.; Li, Z.; Zhuang, J.; Yang, J.; Li, X.; Meng, L.; Zhang, H.; Dong, Y. An Efficient User-Friendly Integration Tool for Landslide Susceptibility Mapping Based on Support Vector Machines: SVM-LSM Toolbox. *Remote Sens.* **2022**, *14*, 3408. <https://doi.org/10.3390/rs14143408>

Academic Editors: Mirko Francioni, Stefano Morelli and Veronica Pazzi

Received: 3 June 2022

Accepted: 14 July 2022

Published: 15 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The occurrence of landslide disasters causes great losses to the economy and human life all over the world every year [1,2]. Natural events such as rainfall [3,4], earthquakes [5,6] and floods [7] often lead to a series of landslides. Landslide susceptibility mapping (LSM) is used to determine the probability of future landslides in the study area by comprehensively analyzing various topographic, geological and hydrological factors, as well as

human activity, alongside historical landslide activity in the study area [8,9]. LSM is of great significance to landslide risk management, human life safety and urban future planning.

In recent years, LSM has attracted the attention of many scholars, and various related articles have been published. The methods of generating landslide susceptibility mapping mainly include empirical modeling based on expert experience [10,11], physically based models [12], data-driven statistical modeling [13–15] and machine learning models [16–19]. Compared with traditional methods, the machine learning models do not rely on expert experience, which reduce the subjectivity of evaluation and generally have high accuracy. With the development of geographic information system (GIS) software and open-source machine learning libraries, the machine learning methods are becoming increasingly popular.

Compared with other machine learning algorithms, the support vector machine (SVM) method has been widely used in calculating landslide susceptibility because of its advantages in solving small-sample, nonlinear and high-dimensional classification problems [5,8]. However, the process of landslide susceptibility assessment using SVM is complicated, consisting of multiple steps such as data preprocessing, influencing factor selection, dataset production, model training and prediction. Generally, when using SVM to generate LSM, researchers must work with a cross-platform. Terrain factors based on the Digital Elevation Model (DEM) (e.g., slope, aspect) rely on platforms such as ArcGIS or QGIS. Model training and parameter optimization usually adopt widely used programming languages such as Python, R or MATLAB. In addition, Excel, SPSS software or programming languages have been used for model accuracy evaluation and statistical analysis in most previous studies.

Tools related to landslide susceptibility mapping are usually available in the form of academic code, which requires users to have programming skills. Some studies have proposed and applied several tools to evaluate landslide susceptibility. Osna et al. [20] developed an independent application (*GeoFIS*) to generate landslide susceptibility maps using the Mamdani fuzzy inference system (FIS). Sezer et al. [11] developed an LSM module based on expert experience with *NetCAD* architecture software. Jebur et al. [21] created a landslide susceptibility mapping toolbox using bivariate statistical analysis (BSA) based on ArcGIS. Zhang et al. [15] provided a landslide susceptibility assessment tool based on the optimized frequency ratio method, which itself is based on the ArcGIS platform. Torizin et al. [22] provided an independent landslide susceptibility assessment application written in Python, Project Manager Suite (LSAT PM). Bragagnolo et al. [23] developed a free and open-source plug-in, namely *r.landslide*, based on the GRASS software of open-source GIS, to generate landslide susceptibility mapping based on an artificial neural network. Sahin et al. [24] integrated R and ArcGIS software and developed a landslide susceptibility mapping toolkit (LSM tool pack) based on logistic regression and random forest. Guo et al. [25] introduced a Python QGIS plugin [26] named FSLAM, which allows us to compute regional shallow landslide susceptibility based on the effective antecedent water recharge and the event rainfall.

Most of the above toolboxes are based on expert experience models or statistical models, such as the weight of evidence method, frequency ratio method and so on. These methods are simple in principle and easy to implement, but with limited accuracy. To date, only a limited number of previous studies have involved the development of landslide susceptibility mapping tools based on machine learning methods. At the same time, most tools only involve model training and prediction, instead of the whole process of LSM. In addition, most studies only use the single-factor pixel value corresponding to landslide point locations as samples for model training. However, landslides usually occur within a region and are affected by characteristics from the surrounding environment. Therefore, problems exist when constructing samples based on a single pixel [27,28]. The realization of regional-scale data construction is often complicated and time-consuming.

To solve the above-mentioned problems, this research develops an LSM toolbox based on the ArcGIS platform (SVM-LSM toolbox). The toolbox includes data preprocessing, factor selection, SVM model training and evaluation, and landslide susceptibility

map prediction, involving the whole process of LSM. Moreover, this toolbox only uses the ArcGIS platform, which avoids cross-platform operation and reduces user input parameters as much as possible. The operation is simple, convenient and user-friendly. The susceptibility prediction process based on sliding windows is time-consuming. This tool provides a multiprocessing rapid prediction tool to sufficiently improve the production efficiency of landslide susceptibility mapping. In addition, a tool for the rapid production of multi-channel block datasets is constructed to improve the efficiency of dataset making. It is worth noting that this toolbox is not limited to the mapping of landslide susceptibility based on SVM and can also be used for other binary classification problems based on SVM. Section 2 of this paper introduces the basic functions of the toolbox and a description of each module; Section 3 discusses the experimental research on the landslide susceptibility mapping of the toolbox in Wuqi County, Shaanxi Province, China, and provides an analysis of the relevant results; and Section 4 presents the conclusion.

## 2. LSM Toolbox

### 2.1. LSM Workflow

An overall flow chart of LSM based on SVM is shown in Figure 1. The process of generating LSM based on SVM consists of data collection, data preprocessing, dataset making, feature selection, model training and susceptibility map prediction. The data collection includes historical landslide data, the coverage of the study area and landslide influencing factors, such as roads, rivers, faults, Normalized Difference Vegetation Index (NDVI), DEM, lithology and rainfall. Among them, landslide points, the coverage of the study area, roads, rivers and faults are vector data, NDVI, DEM and lithology are grid data, and rainfall is the NetCDF-4 (*nc4*) format. Data preprocessing includes calculating topographic factors (such as slope, aspect, etc.) based on DEM, converting line vector data to continuous raster factors, and *nc4* data processing. For raster data, it is also necessary to clip them to the same study area range. Subsequently, based on landslide points and the range of the study area, the same number of non-landslide points are randomly selected to construct negative samples. Then, the dataset is randomly divided into training samples and test samples in the ratio of 7:3. In addition, the Pearson correlation coefficient (PCC) and information gain ratio (IGR) are calculated for all the samples. Influencing factors are selected based on the calculation results; factors with high correlations or with less importance to landslide occurrence are removed. Then, the training and test sets are reconstructed according to the results of the feature selection. Finally, the training set is used to train the model, and an optimal SVM model is obtained through the comprehensive analysis of parameters and evaluation indicators such as accuracy, precision, recall, F1 value, receiver operating characteristics (ROC) and area under the curve (AUC). The optimal model is finally used to predict the susceptibility index of the study area and generate a susceptibility map of the study area for subsequent analysis.

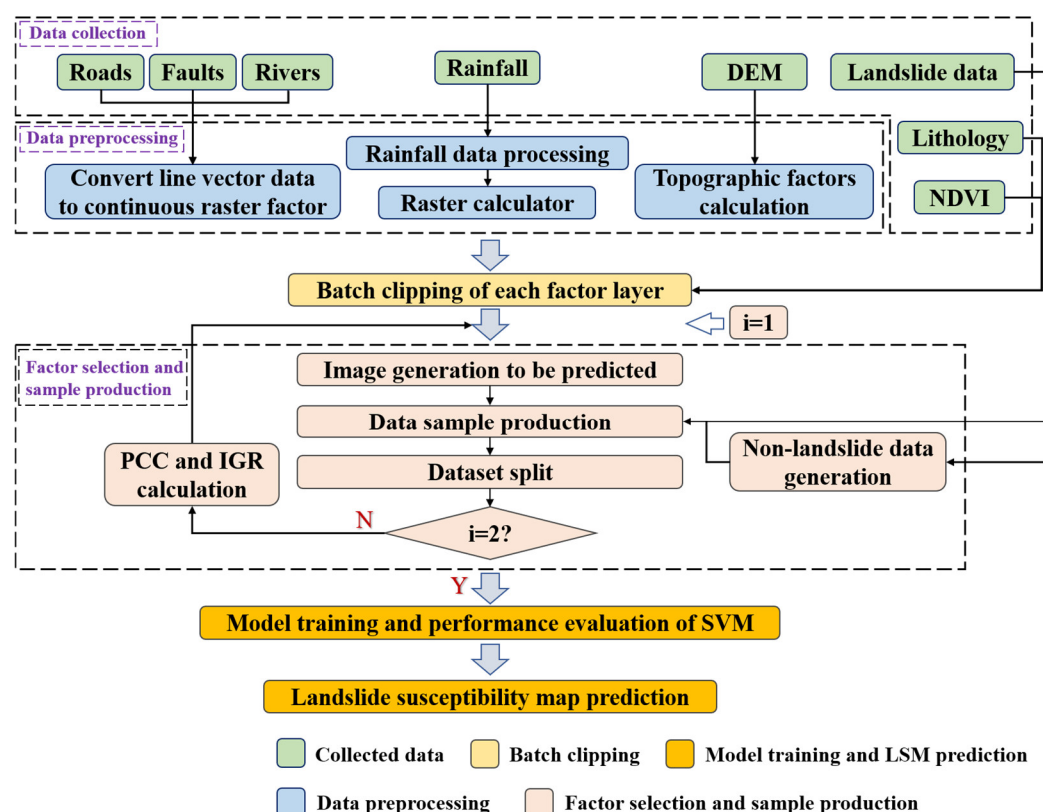


Figure 1. Flowchart of SVM-LSM toolbox.

In this paper, a toolbox is presented to generate landslide susceptibility maps according to the above-mentioned workflow. The LSM toolbox includes three sub-toolboxes: “1 influence factor production”, “2 factor selection and dataset production” and “3 model training and prediction”, as shown in Figure 2. This toolbox is developed based on *ArcPy* and Python language and can be directly integrated into ArcGIS 10.1 (or higher) or ArcGIS Pro software. It is efficient and user-friendly.

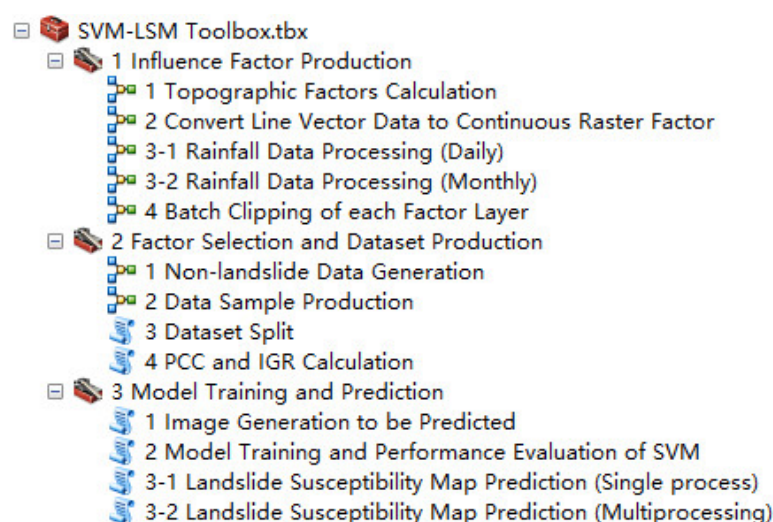


Figure 2. Overall module of SVM-LSM toolbox.

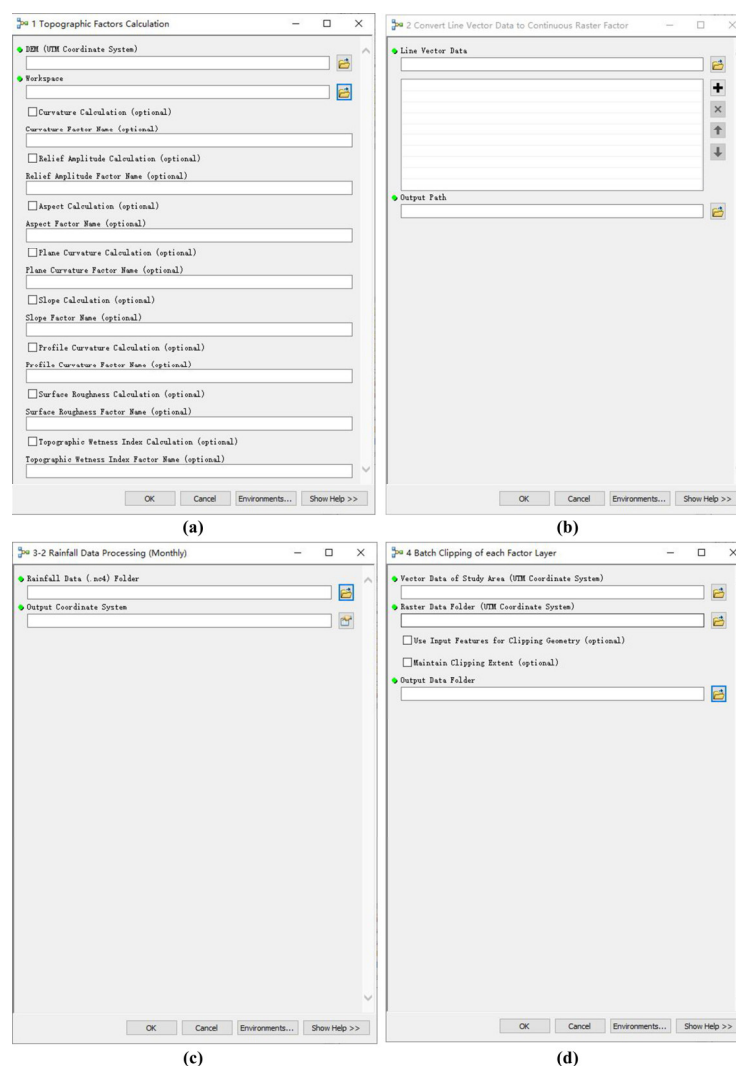
## 2.2. Influencing Factor Production

Landslide influencing factors are various factors that affect the occurrence of landslides through the study of the occurrence mechanism of landslides in the study area. The

occurrence of landslides is affected by various influencing factors. At present, there is no unified standard for the selection of influencing factors. Pourghasemi et al. [29] conducted a statistical analysis on the influencing factors used in the study and found that topographic factors, geological factors and human activities are the most commonly used factors for landslide occurrence. This toolbox provides a tool for generating relevant topographic factors based on DEM, a tool for converting roads, faults and rivers into continuous raster data, and a rainfall processing tool.

### 2.2.1. Topographic Factor Calculation

This tool integrates other factors calculated by DEM, and automatically calculates other topographic factors such as slope, aspect, curvature, plane curvature, profile curvature, relief amplitude, surface roughness, topographic wetness index (TWI) and other topographic factors based on DEM data in the study area. As shown in Figure 3a, it is necessary to only input DEM data and select the factors that need to be calculated. These factors can be calculated optionally according to the needs of users by checking the box in front of the factors to be calculated, but aspect must be calculated when calculating plane curvature, and slope must be calculated when calculating profile curvature, surface roughness or TWI.



**Figure 3.** Influencing factor production toolbox interface: (a) topographic factors calculation; (b) convert line vector data to continuous raster factor; (c) rainfall data processing; and (d) batch clipping of each factor layer.

### 2.2.2. Convert Line Vector Data to Continuous Raster Factor

This tool automatically converts the line vector data of the study area into continuous raster data, such as distance to roads, distance to faults and distance to rivers. The conversion principle adopts Euclidean distance. As shown in Figure 3b, the user only needs to input the line vector data to be converted and the result save path.

### 2.2.3. Rainfall Data Processing

The National Aeronautics and Space Administration (NASA, <https://gpm.nasa.gov/>, accessed on 24 December 2020) provides a Global Precipitation Measurement Mission (GPM). These are high-precision precipitation data obtained using multi-sensors, multi-satellites and multi-algorithms combined with the satellite network and rainfall gauge inversion, with a spatial and temporal resolution up to 0.5 h,  $0.1^\circ \times 0.1^\circ$  [30]. The monthly or daily rainfall data downloaded from NASA are in the *.nc4* format, which is time-consuming and laborious to convert into raster data one by one. Therefore, this tool provides a rainfall batch conversion tool to convert the *.nc4* format data to the *.tif* format raster data. As shown in Figure 3c, the user only needs to input the rainfall data and specify the raster data output coordinate system.

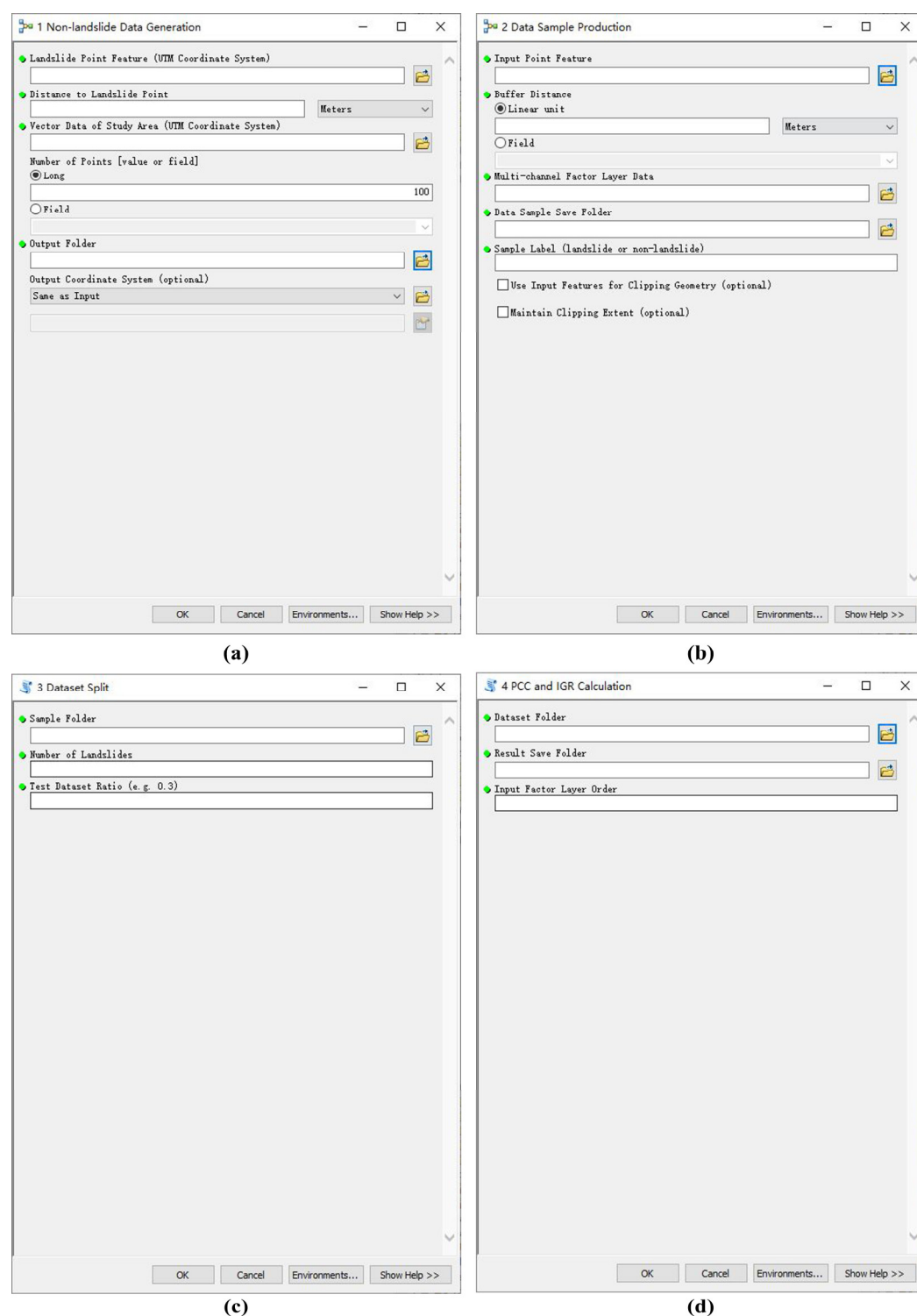
### 2.2.4. Batch Clipping of Each Factor Layer

After the production of the factor layer data, the row and column numbers and coverage of each factor layer data are usually inconsistent. This tool is used to batch clip the raster data of each factor layer according to the vector data of the study area in order to obtain the factor layer data of the study area. As shown in Figure 3d, this tool only needs the user to set the folder where the raster factors are located and the vector data of the study area; it can automatically iteratively select the *.tif* format files for clipping. All the raster data resolutions should be consistent.

## 2.3. Factor Selection and Dataset Production

### 2.3.1. Non-Landslide Data Generation

This tool is used to generate non-landslide point data within the study area vector data layer. As shown in Figure 4a, the user inputs landslide points and the study area vector file and specifies the number of non-landslide points to be selected outside of a buffer and how many meters from the landslide point. First, the tool generates a buffer zone at a specified distance from the landslide point and erases the buffer zone layer on the study area layer to obtain the selectable range of non-landslide sample points. It then uses random points to generate the same number of non-landslide points within the optional range. The non-landslide points should be selected as far from landslide points as possible.



**Figure 4.** Dataset production and factor selected toolbox interface: (a) non-landslide data generation; (b) data sample production; (c) dataset split; and (d) PCC and IGR calculation.

### 2.3.2. Data Sample Production

This tool is used to generate multi-channel block sample raster data from vector point data. As shown in Figure 4b, the user inputs the vector point elements and the multi-channel factor layer data and specifies the buffer distance, which is half of the actual distance represented by the cropped raster size. The tool uses vector point data to create a buffer and iteratively selects the buffer range corresponding to each point vector in order to cut the multi-channel raster data one by one, resulting in a single multi-channel block

dataset of each vector point named after the “FID” attribute value. When the buffer distance is less than the resolution of the raster data, the obtained sample has reached the point at which the landslide point is located.

### 2.3.3. Dataset Split

When using the machine learning methods for model training, it is common to split the samples into a training set and a test set in a certain ratio. The training set is used to train the model and the test set is used to test the generalization of the model and prevent overfitting. As shown in Figure 4c, users can specify the ratio of the training and test sets by themselves. Generally, the ratio of the training and test sets is 7:3. Finally, the sample paths and labels of the training and test sets will be given, respectively (0 for non-landslide and 1 for landslide), and the results are saved in a *txt* file.

### 2.3.4. PCC and IGR Calculation

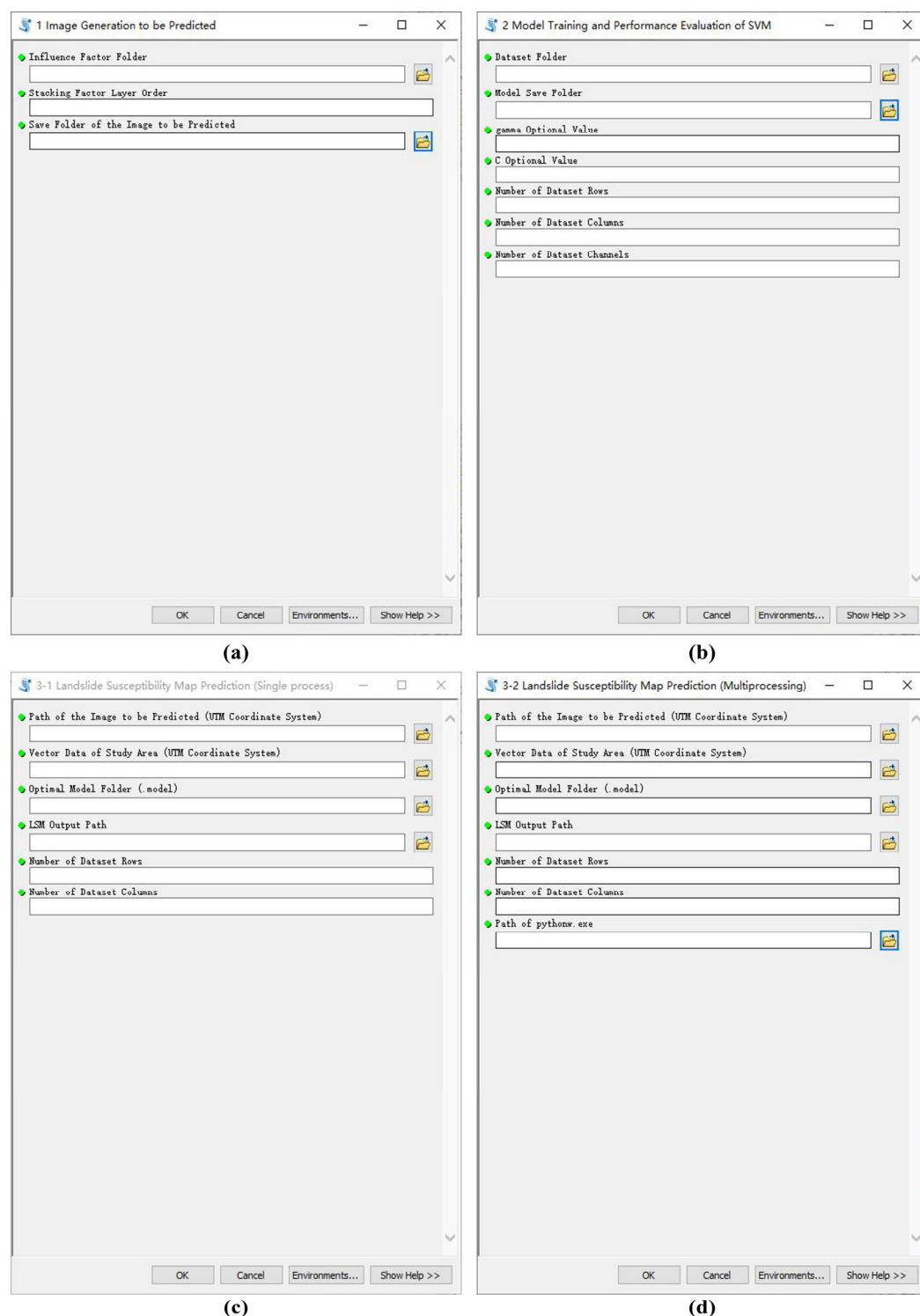
Determining the most effective combination of the influencing factors for landslide susceptibility mapping is of great importance. If the influencing factors are not evaluated, this will not only cause data redundancy but will also affect the execution efficiency and prediction accuracy of the model [31]. At present, there is no optimal solution for the selection of influencing factors, but they typically consist of two parts: correlation analysis and importance evaluation. This toolbox provides two of the most commonly used influencing factor selection methods: PCC and IGR. The PCC is an index used to measure the correlation between the influencing factors. The closer its absolute value is to 1, the stronger the correlation between the two factors. The information gain ratio is an index used to evaluate the importance of each factor layer on landslide occurrence. The higher the IGR value, the greater the impact of this factor on landslide occurrence. Any factor with zero IGR does not influence landslide occurrence. As shown in Figure 4d, this tool calculates PCC and IGR based on the generated data samples and saves them in a *txt* file. Upon comprehensively considering the calculation results, factors with strong correlation and little influence on landslide occurrence are eliminated based on the principle that the lower the correlation is, the greater the importance is.

## 2.4. Model Training and Prediction

### 2.4.1. Image Generation to Be Predicted

The different factor layers are stacked in a certain order to form multi-channel raster data, which is the image to be predicted. It is used for sample production and susceptibility map prediction. As shown in Figure 5a, this tool only requires the input of the path and stacking order of each factor layer. Here, the stacking order of the factor layers used for the image to be predicted should be consistent with the order of the factor layers in the model training samples.





**Figure 5.** Model training and prediction toolbox interface: (a) image generation to be predicted; (b) model training and performance evaluation of SVM; (c) landslide susceptibility map prediction (single process); and (d) landslide susceptibility map prediction (multiprocessing).

#### 2.4.2. Model Training and Performance Evaluation of SVM

This tool is used to generate SVM models with given parameters and provide evaluation results of the accuracy of each model. As shown in Figure 5b, the user enters the directory in which the dataset sample is located along with the number of rows, columns and channels of the dataset. At the same time, the optional values of parameter *gamma*

and penalty factor  $C$  to be adjusted also should be given. The parameter adjustment method used in this tool is the grid search algorithm.

SVM has certain advantages in solving the problem of small-sample classification [32]. The kernel function and slack variable are used to deal with the linear indivisibility of the sample data. At the same time, because the classifier is only determined by the support vector, SVM can effectively avoid overfitting. SVM attempts to classify samples by introducing kernel functions that map landslide influencing factors to a high-dimensional feature space, from which it attempts to locate the optimal hyperplane with the maximum spacing between landslides and non-landslides from the feature space [33]. Xu et al. [5] discussed the influence of different kernel functions of SVM on landslide susceptibility mapping. The results show that the prediction effect of the radial basis function (RBF) in SVM is optimum. Therefore, the kernel function of this tool defaults to RBF.

The susceptibility map is equivalent to a binary classification problem. Landslides are marked as “1” and non-landslides marked as “0”. Thus, the confusion matrix can be constructed according to different combinations of real value and predicted value, and the model accuracy evaluation index can be constructed based on the confusion matrix. In this tool, accuracy, precision, recall,  $F1$  value, receiver operating characteristic (ROC) and area under curve (AUC) were used to evaluate the prediction ability of the model. The calculation formula [6] is as follows.

$$\text{accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ value} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

If the real result and predicted result are landslide, it is called true positive ( $TP$ ); if the real result and predicted result are non-landslide, it is called true negative ( $TN$ ); if the real result is landslide and the predicted result is non-landslide, it is called false negative ( $FN$ ); if the real result is non-landslide and the predicted result is landslide, it is called false positive ( $FP$ ).

In the ROC, false-positive rate ( $FPR$ ) is the x-axis and true-positive rate ( $TPR$ ) is the y-axis. At the same time, the area under the ROC (AUC) is used to quantitatively evaluate the prediction accuracy of methods. The AUC value range is [0,1]. The larger the AUC value, the higher the accuracy of the model classification and the better the accuracy.

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

$$TPR = \frac{TP}{FN + TP} \quad (6)$$

#### 2.4.3. Landslide Susceptibility Map Prediction

This tool is used to predict landslide susceptibility in the study area, based on the optimal model, and obtain the landslide susceptibility map in the study area. In this tool, a sliding window with the same row and column numbers as the dataset is constructed to select the data to be predicted for input into the optimal model to obtain the susceptibility index until all rows and columns are sliding. The tool provides two options: single process (Figure 5c) and multiprocessing (Figure 5d). Single-process and multiprocessing tools can be used under ArcGIS and ArcGIS Pro, but the single-process tool speed is slow and the

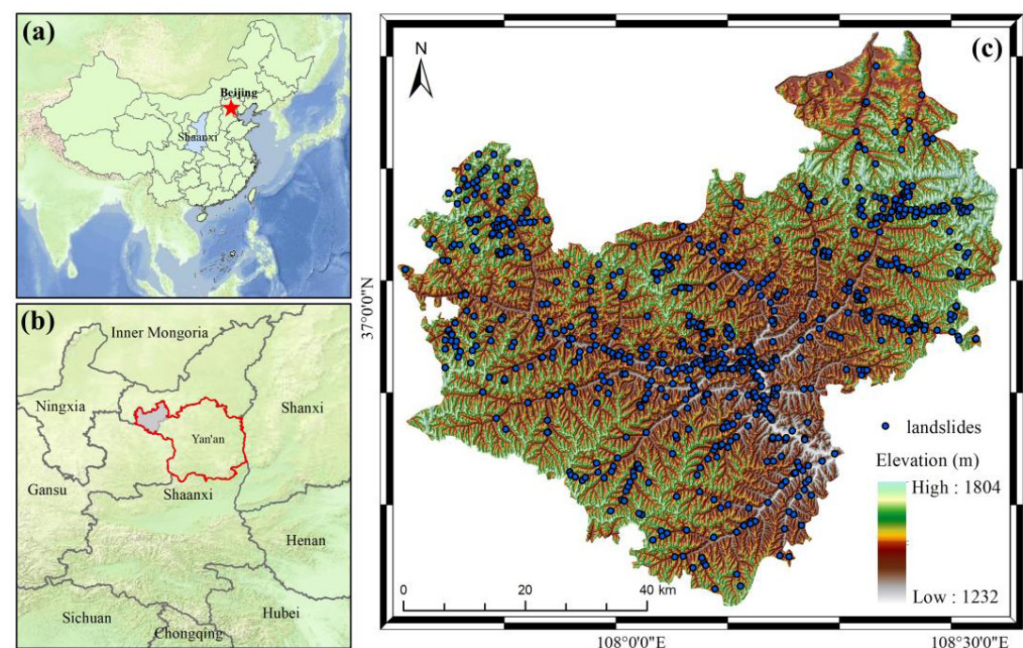
multiprocessing tool is fast. In a single process, the user must only give the image to be predicted, the optimal model and the number of rows and columns of the dataset. In multiprocessing, the user must also specify “pythonw.exe” location.

### 3. Results

Taking Wuqi County, Shaanxi Province, China as an example, the developed toolbox was applied to carry out a landslide susceptibility assessment.

#### 3.1. Study Area

The study area is located in Wuqi County, Yan'an City, Shaanxi Province ( $107^{\circ}38'57''\text{E}$ – $108^{\circ}32'49''\text{E}$ ,  $36^{\circ}33'33''\text{N}$ – $37^{\circ}24'27''\text{N}$ ). It covers a total area of 3791.5 km<sup>2</sup>, encompasses a total population of 145,700 and has an altitude of 1233–1809 m. The study area has a warm, temperate, continental, semi-arid climate. It is dry and windy in spring, sees alternating drought and flood conditions in summer, is cool and wet in autumn and is cold and dry in winter, and the annual average temperature is 7.8 °C. The average annual rainfall is 483.4 mm, and the total coverage of forest and grass is 49.6%. The Wuding and Beiluo River systems lie within the study area. The landform belongs to the hilly and gully area of the Loess Plateau. The terrain fluctuates greatly, the gully is long and the slope is steep [34]. The landslide type in the study area mainly belongs to Loess landslides. During the flood season, rainstorms or continuous rainfall will often induce landslides, collapses and debris flow of different scales, seriously threatening the lives and property safety of local people. Therefore, it is of great practical significance to carry out landslide susceptibility evaluation in Wuqi County. The location of the study area is shown in Figure 6.

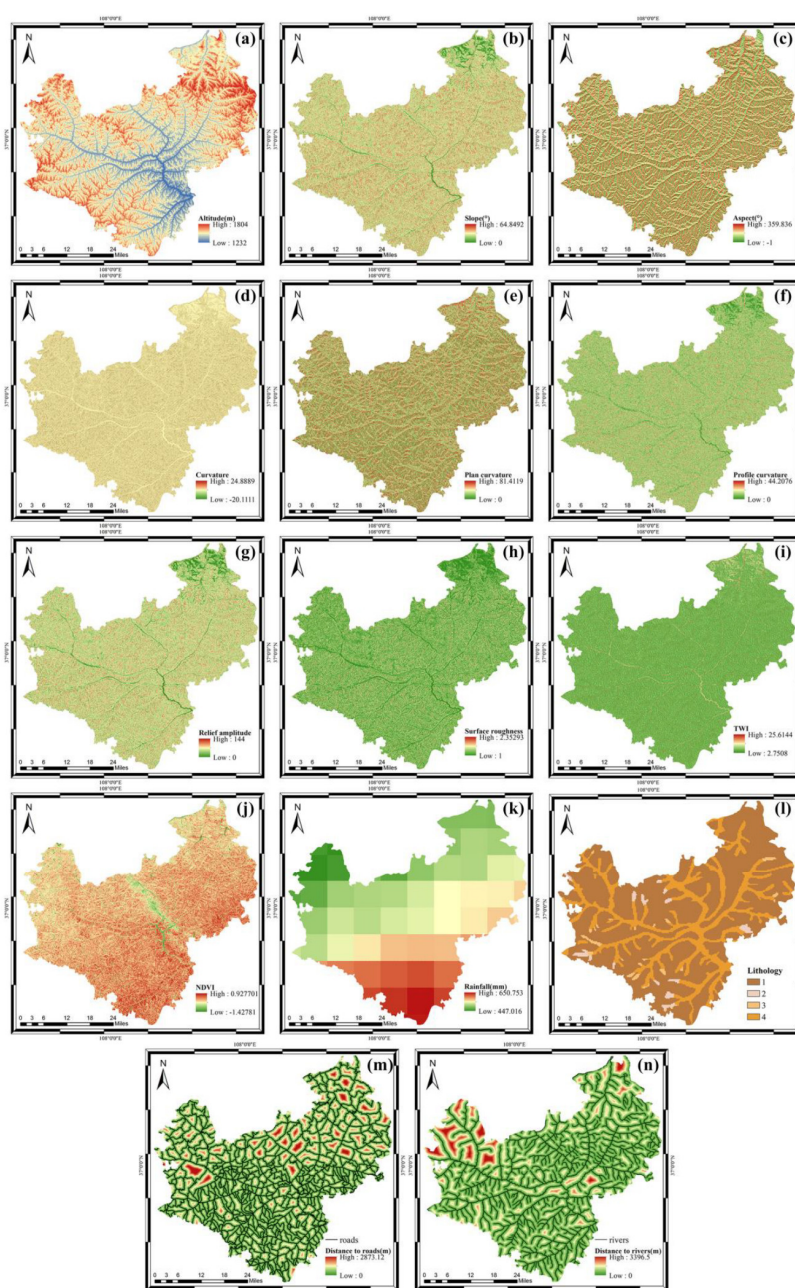


**Figure 6.** (a) Location of Shaanxi Province, (b) location of Wuqi County, Yan'an City, (c) landslide inventory mapping in Wuqi County.

#### 3.2. Preprocessing of Influencing Factors

The influence factor data sources used in this example include DEM, roads, rivers, lithology, NDVI and rainfall. Lithology and NDVI were pre-processed into 30 m resolution raster data. For the acquired DEM data, the “topographic factor calculation” tool is used to generate slope, aspect, curvature, plane curvature, profile curvature, relief amplitude, surface roughness and a topographic wetness index (TWI). At the same time, the

“convert line vector data to continuous raster factor” tool is used to produce the distance to rivers and distance to roads. Since there is no active fault in the study area and it is not affected by active faults, the distance to the fault is not considered. For the rainfall data (.nc4), the “rainfall data processing” tool is used to convert the monthly rainfall data obtained by NASA into the corresponding raster data in batches, and the raster calculator is used to accumulate monthly rainfall data in order to obtain annual rainfall data. Finally, the “batch clipping of each factor layer” tool is used to batch cut the generated influencing factor data according to the vector data of the study area. Finally, a total of 14 landslide influencing factors are generated (Figure 7), and the spatial resolutions of all the factor data are 30 m.



**Figure 7.** Landslide influencing factors in Wuqi County. (a) Altitude, (b) slope, (c) aspect, (d) curvature, (e) plane curvature, (f) profile curvature, (g) relief amplitude, (h) surface roughness, (i) topographic wetness index (TWI), (j) normalized difference vegetation index (NDVI), (k) rainfall, (l) lithology, (m) distance to roads, (n) distance to rivers.

### 3.3. Factor Selection and Sample Generation

There are 789 historical landslides in the study area, which can be divided into 175 large landslides, 417 medium landslides and 197 small landslides. In this study, all the landslide locations are used to construct the landslide dataset. Based on the landslide point data, the “non-landslide data generation” tool was used to randomly generate the same number of non-landslide points, each of which should be at least 1 km away from all of the landslide points in the study area.

Since the calculation of IGR and PCC must be based on all the sample data, the dataset needs to be created before the selection of influencing factors. Firstly, the “image generation to be predicted” tool is used to stack the generated data of 14 influencing factors in the study area in multiple channels. Then, the “data sample production” tool is used to make landslide and non-landslide block datasets based on the superimposed multi-channel images. In addition, the “dataset split” tool is used to divide the training samples and test samples in the ratio of 7:3, before saving the path and labels of the samples to the corresponding *txt* file, respectively. Finally, all the block datasets have fourteen channels, eight rows and eight columns. There are 1104 images in the training set and 474 images in the test set, in which the landslide dataset is marked as 1 and the non-landslide dataset is marked as 0.

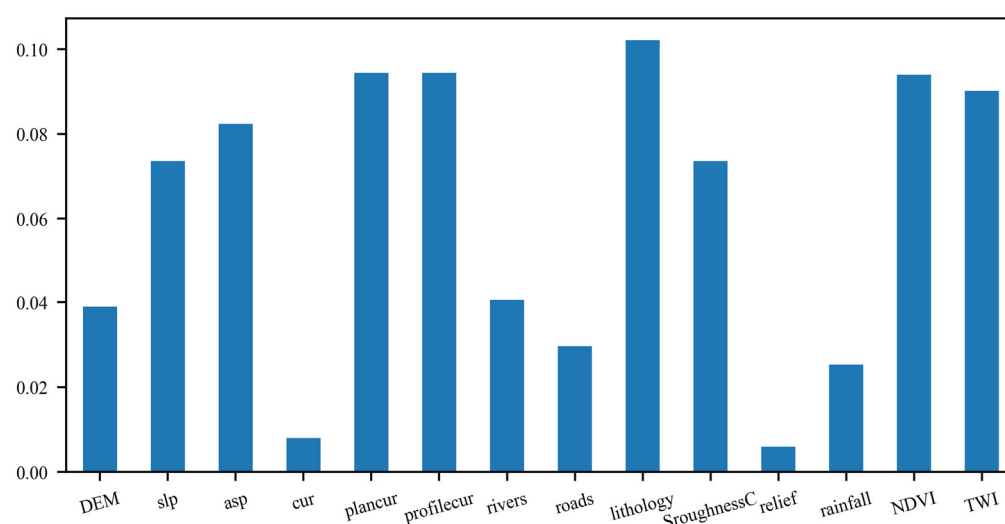
After using the “PCC and IGR calculation” tool to calculate the PCC and information gain ratio of each factor layer based on the data samples, Figure 8 shows the results of the PCC calculation. It can be seen that the correlation coefficients between plane curvature and slope, TWI and slope, and relief amplitude and surface roughness are greater than 0.5. The study area is located in the hinterland of the Loess Plateau which is a typical hilly and gully landscape with high topographic fragmentation and loose soils. The reason for such strong correlations is that the study area often suffers from severe rainfall erosion and river erosion. On the one hand, the greater the slope, the more severe the soil erosion. Therefore, the more complex the surface morphology, the greater roughness and relief amplitude of the surface. On the other hand, the steep slopes with low water retention capacity lead to low soil water content (TWI), and vice versa. Figure 9 presents the calculation results of the information gain ratio. The IGR values of 14 landslide influencing factors are greater than 0, indicating that these factors have an impact on the occurrence of landslides in the corresponding areas. In this study area, lithology has the greatest impact on landslide occurrence, followed by NDVI, plane curvature, profile curvature and TWI, while curvature and relief amplitude have the least impact. Upon a comprehensive analysis of PCC and IGR, the two influencing factors of slope and relief amplitude were removed for Wuqi County, and the remaining 12 influencing factors were used for subsequent research.

According to the evaluation results, the steps of “image generation to be predicted”, “data sample production” and “dataset split” should be repeated in decreasing order of information gain ratio (i.e., lithology, plane curvature, profile curvature, NDVI, TWI, aspect, surface roughness, distance to rivers, DEM, distance to roads, rainfall and curvature) to obtain the final image and sample data for further prediction. The number of channels of all the block datasets used is 12, and their row and column numbers are both eight in the subsequent analysis.





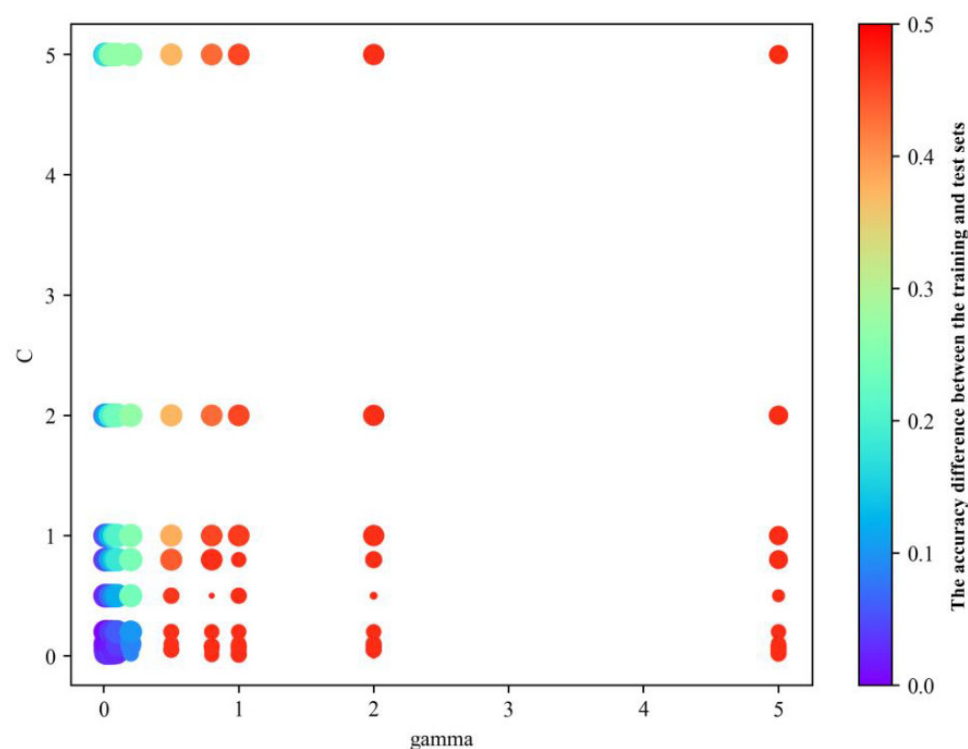
**Figure 8.** Pearson correlation coefficient matrix for the Wuqi County case study. Note that “slp” represents slope, “asp” represents aspect, “cur” represents curvature, “plancur” represents plane curvature, “profilecur” represents profile curvature, “rivers” represents distance to rivers, “roads” represents distance to roads, “lithology” represents lithology, “SroughnessC” represents surface roughness, “relief” represents relief amplitude, and “rainfall” represents rainfall.



**Figure 9.** Information gain ratio for the Wuqi County case study. Note that “slp” represents slope, “asp” represents aspect, “cur” represents curvature, “plancur” represents plane curvature, “profilecur” represents profile curvature, “rivers” represents distance to rivers, “roads” represents distance to roads, “lithology” represents lithology, “SroughnessC” represents surface roughness, “relief” represents relief amplitude, and “rainfall” represents rainfall.

### 3.4. Model Training and Performance Evaluation

The “model training and performance evaluation of SVM” tool is used to train the model based on the generated training data, evaluate the performance with the test set and plot the ROC curve. Of these, the SVM model uses the RBF kernel function. The model has two parameters: *gamma* and penalty factor *C*. The grid search algorithm is used to optimize the parameters, find the optimal set of model parameters and generate the optimal model. The values of parameters *gamma* and *C* are selected from 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.5, 0.8, 1, 2 and 5. Figure 10 shows the AUC values and the difference in accuracy between the training and test sets for different *gamma* and *C* values, which used *gamma* values as horizontal coordinates and *C* values as vertical coordinates. In the figure, the size of the circle represents the AUC value. The larger the circle, the greater the AUC value and the better the model performance. The color of the circle represents the accuracy difference between the training and test sets. If it exceeds 0.5, it is represented by 0.5. The greater the accuracy difference, the higher the degree of overfitting of the model and the worse the generalization performance. Consequently, comprehensive analysis shows that when *gamma* is 0.02 and *C* is 2, the AUC value is high, the accuracy difference is small, and the model is optimal.

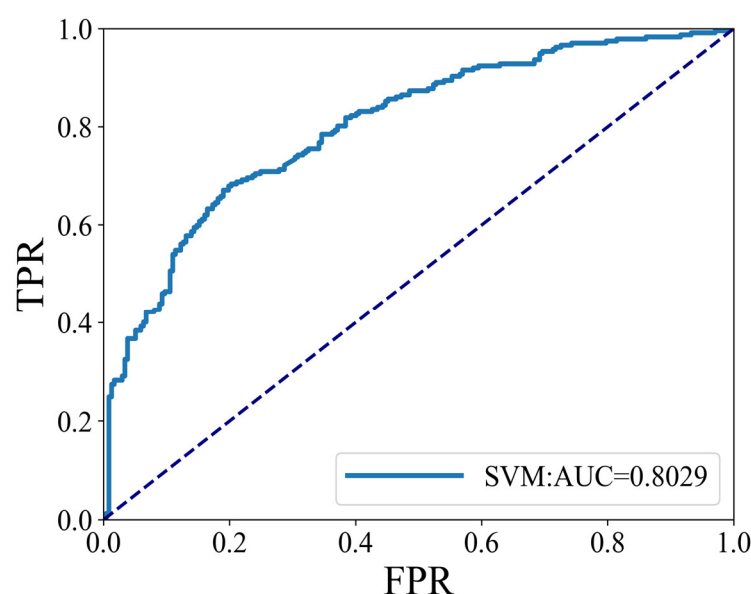


**Figure 10.** AUC values and accuracy differences under different parameter values.

Table 1 shows the performance of the optimal model with the testing dataset, and Figure 11 shows its corresponding ROC curve. Among the 474 testing datasets, 169 landslides and 171 non-landslides were correctly predicted, while 68 landslides and 66 non-landslides were incorrectly predicted. The correct samples predicted by the model accounted for 71.73% of the total samples, with a precision of 71.55% and a recall rate of 72.15%. At the same time, the AUC value of the model is 0.8029, indicating that the model has good prediction performance and the result of the landslide susceptibility map is reliable.

**Table 1.** Evaluation index of the model performance.

Evaluation Index		Results	
Confusion matrix	Landslide	169	68
	Non-landslide	66	171
Accuracy		0.7173	
Precision		0.7155	
Recall		0.7215	
F1		0.7185	
AUC		0.8029	

**Figure 11.** The ROC curve of the optimal model.

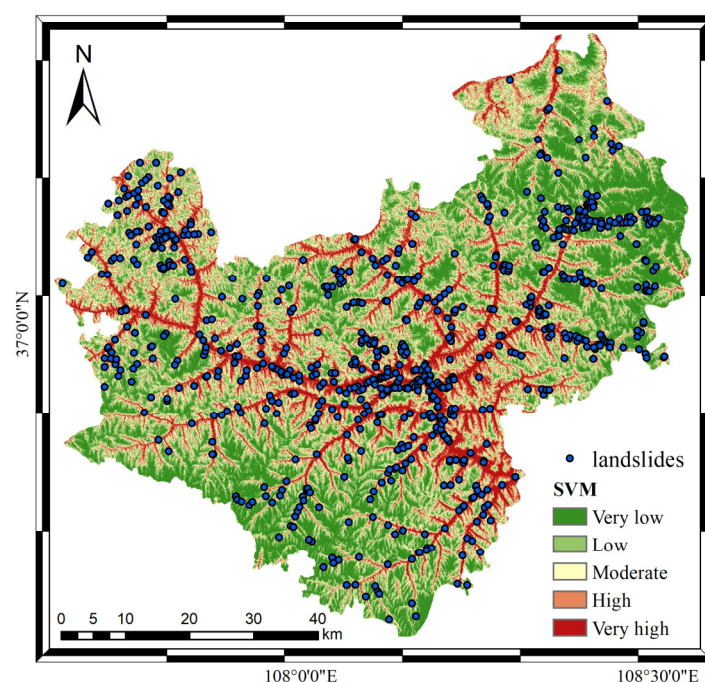
### 3.5. Landslide Susceptibility Map Generation and Analysis

With the trained optimal model, the “landslide susceptibility map prediction” tool is used to predict the generated image unit by unit according to the optimal model. The probability of each evaluation unit being predicted as a landslide is obtained to generate a landslide susceptibility map for the study area. The predicted susceptibility indexes lie between 0 and 1. The larger the susceptibility index is, the more susceptible the area is to landslides. The generated susceptibility map is divided into five levels—very low, low, moderate, high and very high—using the natural break method in ArcGIS. The landslide susceptibility map of Wuqi County after classification is obtained by SVM, as shown in Figure 12.

It is clear in Figure 12 that the areas in Wuqi County with high and very high susceptibility to landslides are mainly concentrated on both sides of rivers severely affected by soil erosion. Low- and very-low-susceptibility areas are mainly distributed in high-altitude areas with limited human activity. The locations of historical landslides are well fitted with the predicted results. The areas where landslides are relatively concentrated are predicted as high and very high susceptibility areas, which is in line with the actual situation. Table 2 shows the proportion of each graded area and the density of landslide points within each grade. It can be seen that the proportion of high- and very-high-susceptibility areas is 29.97%, and the proportion of low- and very-low-susceptibility areas is 49.18%. With increased susceptibility grade, the density of landslide points increases continuously, which is in line with the actual situation of the susceptibility grade. The density of



landslide points in very-high-susceptibility areas is 0.77 and that in very-low-susceptibility areas is 0.04.



**Figure 12.** Classification map of landslide susceptibility in Wuqi County.

**Table 2.** Statistical analysis of each susceptibility class in Wuqi County.

Classes	Area (km <sup>2</sup> )	Proportion (%)	Landslide Density (Number/km <sup>2</sup> )
Very low	924.43	24.28	0.04
Low	948.24	24.90	0.08
Moderate	794.02	20.85	0.14
High	648.42	17.03	0.28
Very high	493.02	12.94	0.77

### 3.6. Toolbox Operation Efficiency Evaluation

Although the “landslide susceptibility map prediction (single process)” and “landslide susceptibility map prediction (multiprocessing)” tools can be used under ArcGIS and ArcGIS Pro, it is recommended that they be used with ArcGIS Pro. Since Python 2.7 installed in ArcGIS is generally 32-bit, it has extremely limited use of memory resources and can only use a maximum of 2G of memory when processing massive data. If it exceeds 2G, a “Memory Error” will appear. Meanwhile, the Python 3 environment used by ArcGIS Pro is 64-bit, which can use more memory than the 32-bit Python, and therefore the “Memory Error” rarely occurs.

Table 3 shows the computation statistics of various tools in ArcGIS and ArcGIS Pro software for Wuqi County, respectively. For evaluation, all the experiments are conducted on a Windows PC ×64 with a 2.30 GHz Gen Intel Core i7-11800H CPU, a 4 GB GeForce RTX 3050 Ti Laptop graphic card and 16 GB of RAM.

**Table 3.** Computation statistics of various tools with different software in Wuqi County.

Tool	ArcGIS	ArcGIS Pro
Topographic factor calculation	58 s	42 s
Convert line vector data to continuous raster factor	1 min 9 s	34 s
Rainfall data processing	57 s	50 s
Batch clipping of each factor layer	18 s	17 s
Non-landslide data generation	2 s	1 s
Data sample production * landslide	5 min 22 s/4 min 46 s	4 min 34 s/4 min 29 s
non-landslide	4 min 56 s/4 min 32 s	4 min 19 s/4 min 15 s
Dataset split *	0.5 s/0.5 s	0.5 s/0.5 s
PCC and IGR calculation	1 min 16 s	57 s
Image generation to be predicted *	3 min 38 s/2 min 45 s	1 min 32 s/1 min 13 s
Model training and performance evaluation of SVM	1 h 55 min 32 s	1 h 8 min 8 s
Landslide susceptibility map prediction (single process)	2 h 53 min 15 s	1 h 26 min 47 s
Landslide susceptibility map prediction (multiprocessing)	21 min 51 s	20 min 12 s
Total †	5 h 19 min 27 s/2 h 48 min 3 s	2 h 58 min 39 s/1 h 52 min 4 s

Notes: “Data sample production”, “dataset split” and “image generation to be predicted” tools must be run twice. \* indicates that the first run time and the second run time, and † shows the total single process running time and the total multiprocessing running time.

As shown in Table 3, the total time of the SVM-LSM toolbox for the ArcGIS single process is 5 h 19 min 27 s and that for the ArcGIS Pro single process is 2 h 58 min 39 s, which improves running efficiency by 44.08%. The main gap in running time is concentrated in the operation of the “susceptibility map prediction” tool. At the same time, the total time of the SVM-LSM toolbox in ArcGIS multiprocessing is 2 h 48 min 3 s and the total time in ArcGIS Pro multiprocessing is 1 h 52 min 4 s, which improves running efficiency by 33.31%. The main difference in the running time is concentrated in the step of the “model training and performance evaluation of SVM”. The abovementioned two differences are mainly due to their difference in the number of bits. Therefore, it is recommended that the toolbox in ArcGIS Pro is run with 64-bit Python. In addition, under the ArcGIS platform, the running time of the “landslide susceptibility map prediction (multiprocessing)” tool is 2 h 48 min 3 s and the running time of the “landslide susceptibility map prediction (single process)” tool is 5 h 19 min 27 s, which shortens running time by nearly 2 h 31 min 24 s and improves running efficiency by 47.39%. Under the ArcGIS Pro platform, the running time of the “landslide susceptibility map prediction (multiprocessing)” tool is 20 min 12 s and the running time of the “landslide susceptibility map prediction (single process)” tool is 1 h 26 min 47 s, which shortens running time by nearly 1 h 6 min 35 s and improves running efficiency by 76.72%. This shows that the multiprocessing prediction tool for the sliding window in this tool can greatly improve the efficiency of susceptibility mapping.

### 3.7. Model Selection: SVM

As mentioned earlier, SVM is used in the toolbox. To assess whether it is optimal to employ SVM, comparisons with two other commonly used models, namely, decision tree (DT) and random forest (RF), are performed. Table 4 shows the operation efficiency and AUC values of different models. The DT model requires two parameters to be adjusted: *max\_depth* and *min\_samples\_leaf*; the RF model requires five parameters to be adjusted: *max\_depth*, *max\_features*, *n\_estimators*, *min\_samples\_leaf* and *min\_samples\_split*; and the SVM model requires two parameters to be adjusted: *gamma* and *C*. For the grid search method, the greater the number of model parameters, the higher the model training time complexity, and the more time-consuming the model tuning is. In terms of model accuracy, for the

same training and testing datasets in Wuqi County, the AUC of the optimal RF model is 0.8372, the AUC of the optimal SVM model is 0.8029, and the AUC of the optimal DT model is 0.7774. The AUC values of SVM and RF model are both higher than 0.8, indicating that these two models can better reflect the landslide susceptibility in this area. Therefore, compared with the three models, the SVM model is friendlier to beginners, with fewer parameters to be adjusted, short running time and high accuracy. Therefore, we choose the SVM model to build the LSM toolbox.

**Table 4.** The operation efficiency and AUC values of different models.

Model	Number of Parameters	Training Time Complexity	LSM Prediction (Multi-processing)	AUC
DT	2	$O(m * n)$	4 min 28 s	0.7774
RF	5	$O(m * n * l * k * j)$	1 h 21 min 25 s	0.8372
SVM (this study)	2	$O(m * n)$	20 min 12 s	0.8029

Notes:  $O$  represent the time complexity;  $m$ ,  $n$ ,  $l$ ,  $k$  and  $j$  represent the number of optional values of different parameters, respectively.

#### 4. Conclusions

This paper develops a tool known as the SVM-LSM toolbox, which integrates the whole process of landslide susceptibility mapping. The toolbox consists of three sub-toolboxes: (1) influence factor production, (2) factor selection and dataset production, and (3) model training and prediction. The tool can be integrated into ArcGIS 10.1 (or higher) as well as ArcGIS Pro. The interface is user-friendly, easy to implement and provides multiprocessing prediction, which greatly improves prediction efficiency. In order to assess the performance of the toolbox, Wuqi County (an area highly prone to Loess landslides) is selected as the study area. Six basic factors are selected and a total of fourteen landslide influencing factors are obtained based on the influencing factor production tool. In the selection of influencing factors, the slope and relief amplitude factors are eliminated according to the results of PCC and IGR. Finally, the model training tool is used to obtain the optimal model according to various evaluation indexes and generate a susceptibility map of the study area.

The results show that the model has good prediction performance and high prediction accuracy. The susceptibility areas of Wuqi County are mainly concentrated along rivers severely affected by soil erosion. In short, the SVM-LSM toolbox optimizes the complex susceptibility mapping process, avoids the cross-platform operation of traditional workflow and greatly shortens the prediction time of the susceptibility map. At present, the toolbox has only been tested with ArcGIS and ArcGIS Pro software on the Windows system. In the future, it will be integrated into other commonly used GIS processing software, such as QGIS, for expansion. Furthermore, more machine learning models can be incorporated, and automatic parameter tuning function can be developed to further improve the user-friendliness and universality of the toolbox.

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

**Funding:** This research was funded by the National Natural Science Foundation of China, grant numbers 41941019 and 42090053; the National Key Research and Development Program of China, grant number 2021YFC3000400; the Shaanxi Province Science and Technology Innovation team, grant number 2021TD-51; the Shaanxi Province Geoscience Big Data and Geohazard Prevention Innovation Team (2022); the Fundamental Research Funds for the Central Universities, CHD, grant

numbers 300102260301, 300102261108, 300102262902, 300102269208 and 300102260404; the Fund Project of Shaanxi Key Laboratory of Land Consolidation, grant number 2019-ZD04.

**Data Availability Statement:** The code of the module is open-source and can be freely downloaded from <https://github.com/HuangWBill/SVM-LSM-Toolbox> (accessed on 2 June 2022). The case data can be applied by emailing [huangwubiao@chd.edu.cn](mailto:huangwubiao@chd.edu.cn).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Koley, B.; Nath, A.; Saraswati, S.; Chatterjee, U.; Bandyopadhyay, K.; Bhatta, B.; Ray, B.C. Assessment of spatial distribution of rain-induced and earthquake-triggered landslides using geospatial techniques along North Sikkim Road Corridor in Sikkim Himalayas, India. *GeoJournal* **2022**, 1–39. <https://doi.org/10.1007/s10708-022-10585-9>.
2. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping landslide susceptibility using data-driven methods. *Sci. Total Environ.* **2017**, *589*, 250–267. <https://doi.org/10.1016/j.scitotenv.2017.02.188>.
3. Camera, C.A.S.; Bajni, G.; Corno, I.; Raffa, M.; Stevenazzi, S.; Apuani, T. Introducing intense rainfall and snowmelt variables to implement a process-related non-stationary shallow landslide susceptibility analysis. *Sci. Total Environ.* **2021**, *786*, 147360. <https://doi.org/10.1016/j.scitotenv.2021.147360>.
4. Qi, T.; Zhao, Y.; Meng, X.; Chen, G.; Dijkstra, T. AI-Based Susceptibility Analysis of Shallow Landslides Induced by Heavy Rainfall in Tianshui, China. *Remote Sens.* **2021**, *13*, 1819. <https://doi.org/10.3390/rs13091819>.
5. Xu, C.; Dai, F.; Xu, X.; Lee, Y.H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. <https://doi.org/10.1016/j.geomorph.2011.12.040>.
6. Yang, X.; Liu, R.; Yang, M.; Chen, J.; Liu, T.; Yang, Y.; Chen, W.; Wang, Y. Incorporating Landslide Spatial Information and Correlated Features among Conditioning Factors for Landslide Susceptibility Mapping. *Remote Sens.* **2021**, *13*, 2166–2190. <https://doi.org/10.3390/rs1312166>.
7. Costache, R.; Ali, S.A.; Parvin, F.; Pham, Q.B.; Arabameri, A.; Nguyen, H.; Crăciun, A.; Anh, D.T. Detection of areas prone to flood-induced landslides risk using certainty factor and its hybridization with FAHP, XGBoost and deep learning neural network. *Geocarto Int.* **2021**, 1–36. <https://doi.org/10.1080/10106049.2021.1973115>.
8. Ma, Z.; Mei, G.; Piccialli, F. Machine learning for landslides prevention: A survey. *Neural Comput. Appl.* **2020**, *33*, 10881–10907. <https://doi.org/10.1007/s00521-020-05529-8>.
9. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* **2018**, *180*, 60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>.
10. Bathrellos, G.D.; Skilodimou, H.D.; Chousianitis, K.; Youssef, A.M.; Pradhan, B. Suitability estimation for urban development using multi-hazard assessment map. *Sci. Total Environ.* **2017**, *575*, 119–134. <https://doi.org/10.1016/j.scitotenv.2016.10.025>.
11. Sezer, E.A.; Nefeslioglu, H.A.; Osna, T. An expert-based landslide susceptibility mapping (LSM) module developed for Netcad Architect Software. *Comput. Geosci.* **2017**, *98*, 26–37. <https://doi.org/10.1016/j.cageo.2016.10.001>.
12. Medina, V.; Hürlimann, M.; Guo, Z.; Lloret, A.; Vaunat, J. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *Catena* **2021**, *201*, 105213. <https://doi.org/10.1016/j.catena.2021.105213>.
13. Chowdhuri, I.; Pal, S.C.; Arabameri, A.; Ngo, P.T.T.; Chakraborty, R.; Malik, S.; Das, B.; Roy, P. Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ. Earth Sci.* **2020**, *79*, 476. <https://doi.org/10.1007/s12665-020-09227-5>.
14. Li, L.; Lan, H.; Guo, C.; Zhang, Y.; Li, Q.; Wu, Y. A modified frequency ratio method for landslide susceptibility assessment. *Landslides* **2017**, *14*, 727–741. <https://doi.org/10.1007/s10346-016-0771-x>.
15. Zhang, Y.-x.; Lan, H.-x.; Li, L.-p.; Wu, Y.-m.; Chen, J.-h.; Tian, N.-m. Optimizing the frequency ratio method for landslide susceptibility assessment: A case study of the Caiyuan Basin in the southeast mountainous area of China. *J. Mt. Sci.* **2020**, *17*, 340–357. <https://doi.org/10.1007/s11629-019-5702-6>.
16. Goyes-Peñafiel, P.; Hernandez-Rojas, A. Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in Popayan, Colombia. *Eng. Geol.* **2021**, *280*, 105958. <https://doi.org/10.1016/j.enggeo.2020.105958>.
17. Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*, 105972. <https://doi.org/10.1016/j.enggeo.2020.105972>.
18. Wang, Y.; Feng, L.; Li, S.; Ren, F.; Du, Q. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *Catena* **2020**, *188*, 104425. <https://doi.org/10.1016/j.catena.2019.104425>.
19. Fang, Z.; Wang, Y.; Duan, G.; Peng, L. Landslide Susceptibility Mapping Using Rotation Forest Ensemble Technique with Different Decision Trees in the Three Gorges Reservoir Area, China. *Remote Sens.* **2021**, *13*, 238. <https://doi.org/10.3390/rs13020238>.
20. Osna, T.; Sezer, E.A.; Akgun, A. GeoFIS: An integrated tool for the assessment of landslide susceptibility. *Comput. Geosci.* **2014**, *66*, 20–30. <https://doi.org/10.1016/j.cageo.2013.12.016>.
21. Jebur, M.N.; Pradhan, B.; Shafri, H.Z.M.; Yusoff, Z.M.; Tehrany, M.S. An integrated user-friendly ArcMAP tool for bivariate statistical modelling in geoscience applications. *Geosci. Model Dev.* **2015**, *8*, 881–891. <https://doi.org/10.5194/gmd-8-881-2015>.

22. Torizin, J.; Schüßler, N.; Fuchs, M. Landslide Susceptibility Assessment Tools v1.0.0b—Project Manager Suite: A new modular toolkit for landslide susceptibility assessment. *Geosci. Model Dev.* **2022**, *15*, 2791–2812. <https://doi.org/10.5194/gmd-15-2791-2022>.
23. Bragagnolo, L.; da Silva, R.V.; Grzybowski, J.M.V. Landslide susceptibility mapping with r Landslide: A free open-source GIS-integrated tool based on Artificial Neural Networks. *Environ. Model. Softw.* **2020**, *123*, 104565. <https://doi.org/10.1016/j.envsoft.2019.104565>.
24. Sahin, E.K.; Colkesen, I.; Acemali, S.S.; Akgun, A.; Aydinoglu, A.C. Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. *Comput. Geosci.* **2020**, *144*, 104592. <https://doi.org/10.1016/j.cageo.2020.104592>.
25. Guo, Z.; Torra, O.; Hürlimann, M.; Abancó, C.; Medina, V. FSLAM: A QGIS plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environ. Model. Softw.* **2022**, *150*, 105354. <https://doi.org/10.1016/j.envsoft.2022.105354>.
26. Bartolini, S.; Cappello, A.; Marti, J.; Negro, C. QVAST: A new Quantum GIS plugin for estimating volcanic susceptibility. *Nat. Hazards Earth Syst. Sci. Discuss.* **2013**, *1*, 4223–4256. <https://doi.org/10.5194/nhessd-1-4223-2013>.
27. Gaidzik, K.; Ramírez-Herrera, M.T. The importance of input data on landslide susceptibility mapping. *Sci. Rep.* **2021**, *11*, 19334. <https://doi.org/10.1038/s41598-021-98830-y>.
28. Regmi, N.R.; Giardino, J.R.; McDonald, E.V.; Vitek, J.D. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. *Landslides* **2013**, *11*, 247–262. <https://doi.org/10.1007/s10346-012-0380-2>.
29. Pourghasemi, H.R.; Yansari, Z.T.; Panagos, P.; Pradhan, B. Analysis and evaluation of landslide susceptibility: A review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arab. J. Geosci.* **2018**, *11*, 193. <https://doi.org/10.1007/s12517-018-3531-5>.
30. Stanley, T.A.; Kirschbaum, D.B.; Benz, G.; Emberson, R.A.; Amatya, P.M.; Medwedeff, W.; Clark, M.K. Data-Driven Landslide Nowcasting at the Global Scale. *Front. Earth Sci.* **2021**, *9*, 378. <https://doi.org/10.3389/feart.2021.640043>.
31. Tien Bui, D.; Ho, T.-C.; Pradhan, B.; Pham, B.-T.; Nhu, V.-H.; Revhaug, I. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks. *Environ. Earth Sci.* **2016**, *75*, 1101–1123. <https://doi.org/10.1007/s12665-016-5919-4>.
32. Awad, M.; Khanna, R. Support Vector Machines for Classification. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Awad, M., Khanna, R., Eds.; Apress: Berkeley, CA, USA, 2015; pp. 39–66.
33. 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. <https://doi.org/10.1016/j.cageo.2012.08.023>.
34. Wang, S.; Zhuang, J.; Zheng, J.; Fan, H.; Kong, J.; Zhan, J. Application of Bayesian Hyperparameter Optimized Random Forest and XGBoost Model for Landslide Susceptibility Mapping. *Front. Earth Sci.* **2021**, *9*, 617. <https://doi.org/10.3389/feart.2021.712240>.