

A SPATIOTEMPORAL FUSION NETWORK TO MULTI SOURCE HETEROGENEOUS DATA FOR LANDSLIDE RECOGNITION

Bolin Li¹, Ling Han^{1,*}, Liangzhi Li²

¹School of Land Engineering, Chang'an University, Xi'an 710064, China - libolin@chd.edu.cn, hanling@chd.edu.cn

²College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710064, China - liliangzhi@chd.edu.cn

KEY WORDS: Landslide, Multi-source heterogeneous data, Spatiotemporal fusion, Recognition, Deep learning.

ABSTRACT:

In recent years, the frequency of landslide disasters has been increasing year by year due to the extension of human activities to the natural environment. Fast and detailed landslide surveys are important for landslide disaster prediction and management. There are many driving factors for landslide formation, and most of the current deep learning-based landslide identification methods use optical remote sensing images in a short period or a few types of fused data for prediction. Therefore the upper limit of accuracy they can achieve is low. This paper proposes a landslide identification network model based on the spatio-temporal fusion of heterogeneous data from multiple sources. The model takes observations such as time-series optical remote sensing images, DEM, geological formations, and meteorological data as inputs. To address the problems of non-uniform data forms and redundancy caused by time-series data, we design the temporal phase fusion module of coupled CNN-LSTM to fuse the temporal features of multi-source data based on the extraction of their spatial features. Subsequently, we design the spatial feature fusion module based on DCNN-DBN to realize the deep expression of temporal phase and spatial features of landslides and improve the recognition efficiency and accuracy of the network. Through experimental verification, the AUC value of our proposed model is 0.8976, the F1 score is 0.8352, and the MIoU is 0.8624. The evaluation results reflect that the model can provide support for large-scale landslide disaster investigation.

1. INTRODUCTION

Landslides are characterized by their large scale, suddenness, and concealment. In recent years, the frequency of landslides has been increasing year by year due to the extension of the scope of human activities to the natural environment (Li and Li, 2012). Therefore, landslides need to be identified using wide-scale landslide identification techniques. The traditional landslide survey method is a mainly on-site survey, which has poor timeliness and the survey results are not comprehensive and objective enough to achieve rapid and accurate identification of landslides on a large spatial scale (Xing and Liu, 2018).

Remote sensing is capable of rapidly acquiring macroscopic information on the surface over a large area and has become a key technology in the field of landslide identification and monitoring (Li et al., 2021). Manual interpretation refers to the extraction of landslide information from remote sensing images and terrain surface features (Petley et al., 2005). This method generally has a high degree of accuracy but is poor in timeliness and objectivity. With the development of computer image processing technology, methods such as image element-based, object-oriented, and machine learning are applied to the field of remote sensing landslide identification, making the recognition process more automated (Jaedicke et al., 2014). When using very high resolution remote sensing images (VHR) for landslide identification, the object-oriented approach has higher identification accuracy than the image-based approach (Keyport et al., 2018). A machine learning approach enables long-time series landslide displacement prediction and quantifies its uncertainty (Jiang et al., 2021). Most of the above automatic or

semi-automatic identification methods are based on the shallow features of landslides on optical remote sensing images to achieve landslide identification and localization. These methods are severely limited by the quality of the data and the upper limit of the recognition accuracy that can be achieved is low (Chen et al., 2021).

With the development of artificial intelligence, deep learning methods have been widely used in the field of landslide identification. Compared with machine learning, the data-driven deep learning method has a larger sample capacity and does not require manual selection and construction of feature layers when processing landslide features, further increasing the degree of recognition automation. Landslide deformation monitoring based on multi-source heterogeneous data can be realized by using BP neural network (Wang et al., 2021). Feature branching network can extract multi-factor features, and combined with a deep learning network model, it can realize automatic landslide identification based on multi-source data fusion (HUANG et al., 2022).

The formation of landslides is influenced by multiple factors, and the formed landslide form is susceptible to large changes with time (Xu et al., 2018). Currently, deep learning recognition methods have started to analyze landslide multi-source observation data. However, the types of data fused at this stage of research are relatively few, and the fusion mechanism is immature, which still cannot express the deep features of landslides. Moreover, few current studies have considered the multi-temporal data fusion problem, which makes it difficult to obtain complete decoding knowledge of landslides.

To address the above problems, this paper proposes a landslide identification model based on the fusion of multi-source het-

* Corresponding author (E-mail address: hanling@chd.edu.cn)

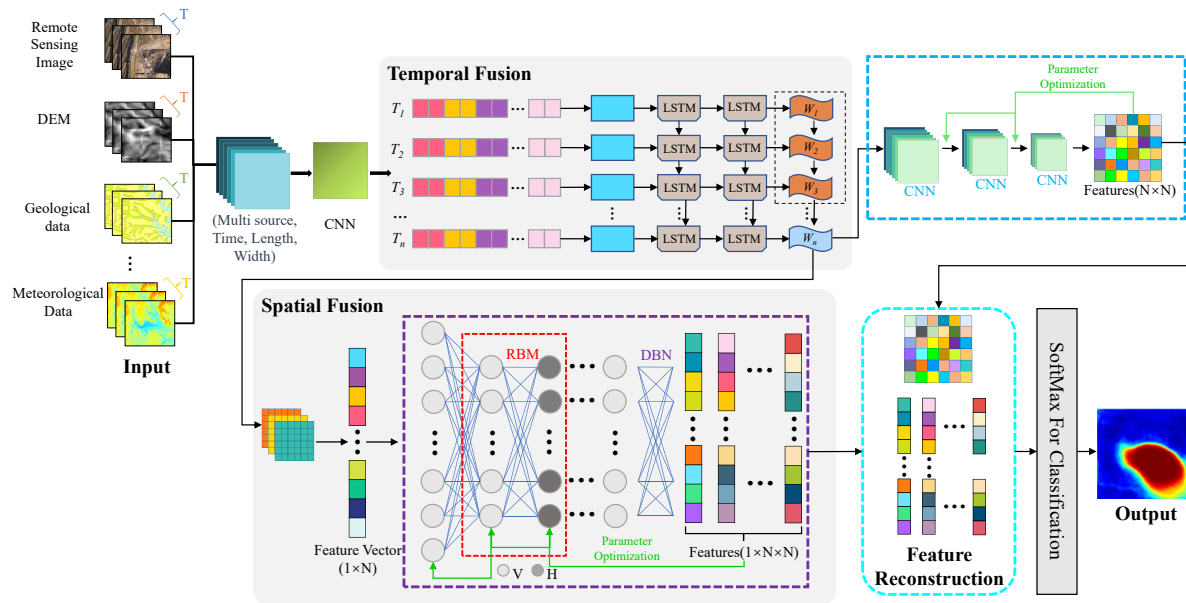


Figure 1. Overview of the proposed method. The network has three components: 1. The CNN performs feature extraction on the 3D convolutional pair (multi-source, temporal, long, and wide) data, and the extracted temporal-spatial features are selectively compressed by the LSTM to achieve data fusion in the temporal dimension, and then the single-dimensional landslide features are restored to their original data dimensions by the upsampling network. 2. The DCNN-DBN network expands all the shallow features of each class of image elements into a one-dimensional vector as the input variables of the DBN. The fused feature data and the original data are used as the input of the DCNN network. 3. After the information extraction of the input data through DBN and DCNN networks respectively, the output feature vectors of both are feature reconstructed and merged into a feature matrix and input to a logistic regression classifier for discrimination, and then landslide identification is achieved.

erogeneous spatio-temporal data. The model has the following features: (1) A temporal phase fusion module based on CNN and LSTM is proposed. While extracting landslide features on time series using CNN, the fusion of landslide temporal features is achieved by LSTM with temporal memory. (2) A spatial feature fusion module based on DCNN and DBN is proposed. DCNN makes up for the lack of spatial representation capability of DBN on large-scale information extraction, expands the input range of image element neighborhood information, and improves data utilization. (3) A feature reconstruction module is designed. The output feature vectors of DCNN and DBN are combined and fed into the logistic regression classifier for discrimination on the one hand and involved in parameter optimization of DCNN and DBN on the other hand. The model finally outputs pixel-level classification results, which are of high practical value for landslide hazard risk assessment and management.

2. THEORY AND METHODOLOGY

2.1 General Framework of the Model

The existing feature fusion methods cannot achieve the high-dimensional fusion representation of landslide features because of the inconsistency in magnitude and spatial scale between time-series remote sensing images and landslide-related environmental factor data (Hu et al., 2016). To fully exploit the information provided in the long time series observation data of various landslide impact factors, we propose a landslide identification network framework with time series remote sensing images, DEM, geological formations, meteorological data, and other observation data as input, as shown in Figure 1. The main

body of the framework consists of three components: a CNN-LSTM-based landslide timing feature fusion module, a DCNN-DBN-based landslide automatic recognition module, and a data feature reconstruction module. The framework is designed to implement two major functions: spatio-temporal feature fusion of multi-source heterogeneous data and landslide identification based on multi-feature fusion data.

In this model, optical remote sensing images and landslide-related environmental factor data of the study area are firstly input into CNN in the form of "multi-source, time-series, long, wide" for feature extraction. Subsequently, the extracted spatial features of the temporal sequence are input to LSTM to compress their temporal features and then recovered to the input data dimension by an up-sampling network to complete the fusion of temporal phase features of multi-source heterogeneous data. The time-phase fused data are input to DCNN and DBN networks respectively for information extraction after dimensional reconstruction, and then the output feature vectors of both are feature reconstructed and merged into a feature matrix, which is input to a logistic regression classifier for discrimination, and finally, landslide recognition is realized.

2.2 CNN-LSTM-Based Landslide Timing Feature Fusion Module

To address the problem that the temporal information is difficult to be fused and utilized, we constructed a landslide temporal phase feature fusion module coupled with CNN and LSTM, as shown in Figure 2. First, the extraction of spatial features on time series is accomplished by using a multiple-input-to-multiple-output CNN network. Then a multiple-input-to-single-output LSTM network is built to compress the

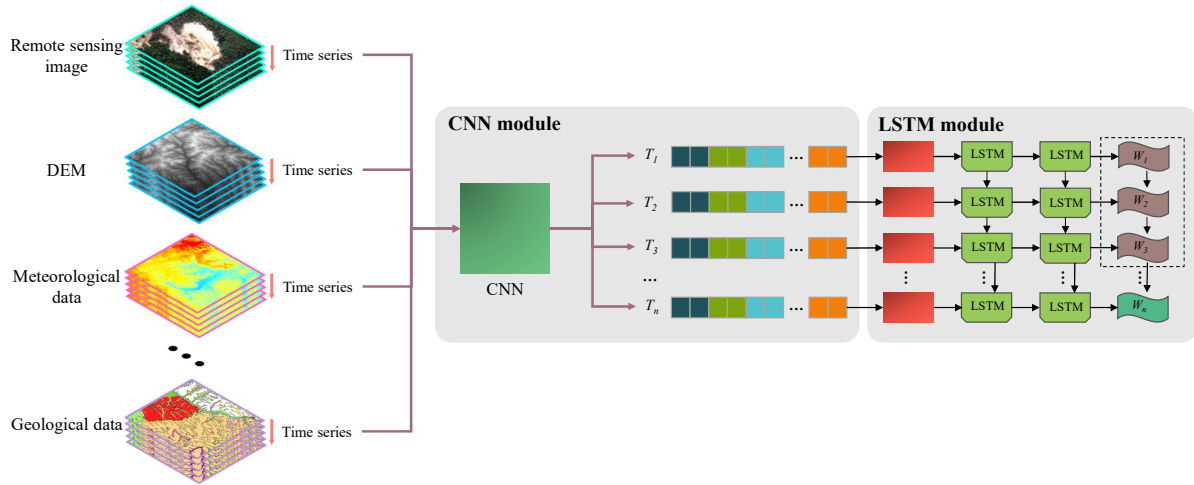


Figure 2. CNN-LSTM module structure diagram.

time-series spatial data and generate a spatio-temporal high-dimensional feature map of the landslide. On this basis, the up-sampling layer is constructed to generate feature maps of the same size as the input data, and the landslide spatial distribution and boundary contour information are effectively maintained based on feature fusion.

2.2.1 Spatial Feature Extraction: In this part, CNN is used to extract landslide spatial features by time series. We construct a convolutional network structure with multiple inputs to multiple outputs and design a feature extraction network layer that takes into account both temporal inputs and outputs. This part finally generates a high-dimensional feature representation of landslide multi-source heterogeneous spatio-temporal data at both temporal and spatial scales.

First, the given temporal multi-source heterogeneous data are combined into a 3D convolutional pair (multi-source, temporal, length, width) and fed into CNN for spatial feature extraction. 3D convolution is very suitable for multidimensional feature extraction (Jia et al., 2018). 2D convolution computes features from the spatial dimension only, while 3D convolution can compute features from both spatial and temporal dimensions, and form multidimensional data by stacking multiple continuous data together with 3D convolution kernels. Formally, the characteristic map of the i th element of the value at x , y , and z at layer j is given by the following equation.

$$v_{ij}^{xyz} = b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} \quad (1)$$

where R_i is the size of the 3D convolution kernel along the time dimension and w_{ijm}^{pqr} is the value of the kernel of (p, q, r) connected to the m th feature map of the previous layer.

Since the 3D convolution kernel weights are replicated throughout the multidimensional data, they should be similar to normal convolution. The number of feature maps is increased by generating multiple types of landslide features from the same set of lower-level feature maps.

2.2.2 Fusion of Temporal Phase Features: Long short-term memory network (LSTM) is a variant of recurrent neural network (RNN), which can handle the long-term dependence of data in time series and solve the problems of long-term dependence as well as gradient disappearance and explosion in the training process of RNN in long time series (Fuzhong et al., 2022). The input of temporal features from multiple sources of data will inevitably cause information redundancy and make it difficult to fit the network, so this part mainly implements the fusion of landslide temporal features. We construct a multi-input to single-output LSTM network with temporal memory based on a many-to-many CNN model. LSTM receives the spatial features of the landslide temporal sequence extracted by CNN, handles the long-term dependency in the temporal data, remembers the temporal landslide feature information for a long time, achieves the compression of the temporal features, and obtains the fusion of temporal and spatial features in the same dimension.

The LSTM consists of three unique "gate" structures (forgetting gate, input gate, and output gate) and a cell for memory storage, as shown in Figure 3. The internal updating process is divided into 3 main steps as follows.

1) Establish forgetting gates for temporal fusion data and input landslide-related factor feature data. The formula for calculating the forgetting gate is as follows.

$$z_f = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

where z_f is forgotten gate control activation value, σ is sigmoid function, W_f is weighting matrix for forgetting gating, h_{t-1} is output value of temporal phase fusion data and factor feature data at the previous moment, x_t is input value of temporal phase fusion data and factor feature data at the current moment, and b_f is forgotten gating bias item.

The forgetting gate will determine which information is discarded from the cell state information C_{t-1} from the previous moment. The gate reads the output value h_{t-1} of the temporal phase fusion data and factor feature data at moment $t-1$, the input value x_t and the forgetting gating bias term b_f at the current moment, and calculates the forgetting gating activation value

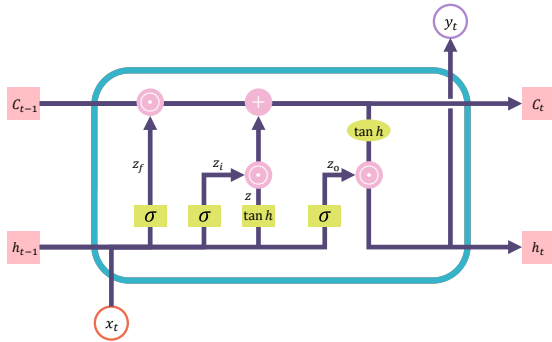


Figure 3. LSTM internal structure diagram.

z_f by the sigmoid function. The output activation values range from 0 to 1 and represent the probability that the previous layer of cell states C_{t-1} is forgotten, with 1 being fully retained to C_t and 0 being fully discarded.

2) Compute the state information C_t of the time-phase fusion data and the factor feature data at moment t . The process begins by determining the information input to the current moment cell state C_t , including the memory gating activation value z_i and the memory cell input state z . The calculation equations are as follows.

$$z_i = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$z = \tanh(W \cdot [h_{t-1}, x_t] + b) \quad (4)$$

$$C_t = z_f \times C_{t-1} + z_i \times z \quad (5)$$

where W_i is memory gating weighting matrix, W is weight matrix of the input states of the memory cell, b_i is bias term for input gating, b is bias item for the input state of the memory unit, \tanh is hyperbolic tangent function, and \odot is hadamard product.

3) Compute the output state h_t of the time-phase fusion data and the factor feature data at the current moment (moment t). z_o is the output gating to control the extent to which the state information C_t at moment t is passed to h_t . The calculation equations are as follows.

$$z_o = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = z_o \times \tanh(C_t) \quad (7)$$

where W_o is weight matrix for output gating and b_o is bias term for output gating.

2.3 DCNN-DBN-Based Landslide Spatial Feature Fusion Module

At present, one of the challenges of landslide recognition based on deep learning is the construction of the network. Most current deep learning landslide recognition networks are for image features on a single high-resolution remote sensing image, and the model structure cannot effectively respond to the complex features of multi-source heterogeneous spatio-temporal data, and the expression ability of multi-feature fusion data is weak, and there are certain misidentification and missed identification (Li et al., 2020).

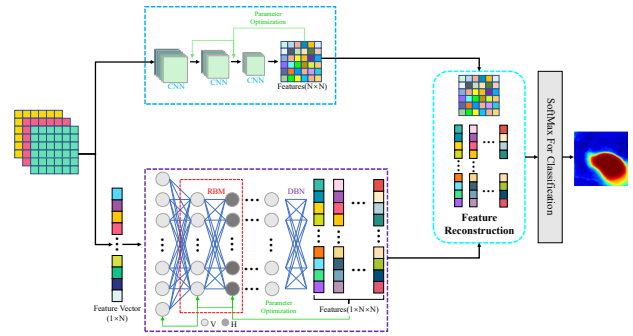


Figure 4. DCNN-DBN module structure diagram.

To address the above problems, we designed a landslide spatial feature fusion module coupling DCNN and DBN, as shown in Figure 4. The joint network expands all the shallow features of each class of image elements into a one-dimensional vector, which is used as the input variable for the DBN. DCNN uses fused feature data and raw data as network inputs. The two networks are trained simultaneously. After the input data are extracted by DBN and DCNN for information extraction, DBN outputs a one-dimensional feature vector and DCNN outputs a two-dimensional feature matrix, and the output feature vectors of both are combined into a new feature matrix and input to the logistic regression classifier for discrimination. The joint model expands the input range of the image element neighborhood information, fills the problem that the perceptual field range needs to be increased after the fusion of multiple features, and improves the diversity of spatial information by increasing the network type and network depth.

The Deep Belief Network (DBN) used in this module is a deep network model consisting of a multilayer unsupervised RBM network and a layer of supervised backpropagation (BP) network, which can effectively handle data classification problems (Wu et al., 2020). The training process of DBN consists of unsupervised pre-training and supervised fine-tuning. The pilot part of DBN, i.e., the RBM layer, is pre-trained layer-by-layer (Layer-wise) to mine the hidden features among the landslide feature data and obtain the high-dimensional representation of landslide features, which is used as the input of the BP layer. The BP layer classifies the feature vector (that is, the extracted features) after RBM training, compares the landslide classification result with the expectation to get the error, and passes the error backward layer by layer to fine-tune the DBN weights.

2.4 Data Feature Reconstruction Module

Model parameter optimization is one of the core problems of deep learning. After feature extraction, DCNN outputs a two-dimensional feature map, and DBN outputs a one-dimensional feature vector, and the output features of both are reconstructed into a two-dimensional feature matrix. In the process of reverse error propagation, the fused landslide feature matrix is split into a one-dimensional feature vector and a separate feature map to participate in parameter optimization in DBN and DCNN, respectively.

The choice of the parameter optimization algorithm is more related to the model type and objective function. Different optimization algorithms have different advantages and disadvantages for model training, and the choice of the algorithm varies from model to model. The project uses DBN and DCNN as the base model, and according to the structural characteristics of the

two networks and the dimensionality of the obtained data, and considering the designed loss function, the Bayesian algorithm and random search are chosen as the parameter optimization algorithms of the model in this paper. A storage structure is designed in parameter passing to pass the parameters after the joint action of stochastic search and Bayesian optimization to the next layer of the network, completing the optimization of each layer of the network.

2.5 Loss Function Design

The role of the loss function is to estimate the closeness between the predicted and true values of the model, and it is a measure of the robustness of the model. For the deep network structure designed in this study, we use the superposition function representation of mean square error loss and cross-entropy loss as the loss function of the network. On the one hand, it is due to the suitability of using mean square error discriminative neurons in feature reconstruction experiments, and on the other hand, it is because landslide feature extraction and identification itself belongs to a semantic segmentation problem, and in the semantic segmentation problem, the cross-entropy loss function can effectively describe the degree of proximity between the true and predicted values. The specific design is as follows.

$$L = \frac{1}{2n} \sum_1^n \|y - a\|^2 - \frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)] \quad (8)$$

$$a = f(w \cdot x + b) \quad (9)$$

where L is loss function value, y is landslide sample actual label value, a is model predicted value, x is input to the model, n is total number of landslide samples, f is activation function, and w, b are network parameters.

Due to the variation of the network structure, the root means square error and the cross entropy superposition function need to be used simultaneously with the SoftMax identification function in the overall calculation process. The overall model prediction, loss acquisition, and learning process are as follows: (i) the discriminant score of the landslide category is obtained from the last layer of the network, and this score is obtained by the joint action of the fused feature discriminant; (ii) the landslide probability is calculated by the SoftMax classification function; (iii) the loss between the model predicted landslide probability and the real landslide category value is calculated by the root to mean square error and cross-entropy loss superposition function.

3. EXPERIMENTAL DESIGN

To verify the rationality and effectiveness of the model proposed in this paper, we trained and tested the model, and set up comparison experiments and ablation experiments. In this study, a personal computer is used as the experimental platform, and the computer configuration information is as follows: CPU model is 11th Gen Intel(R) Core(TM) i7-11700F, with 16.0GB of RAM, graphics card model is NVIDIA GeForce RTX 3060, and OS version is Windows 10 Professional. All network codes used in this paper are written based on PyTorch deep learning architecture with a built-in Python 3.8 development environment.

3.1 Data Set Collection and Processing

In this paper, the Sino-Pakistani Karakorum Highway landslide catalog and feature dataset (Su et al., 2022) and the Sichuan and surrounding landslide mudflow disaster high-precision aerial imagery and interpretation dataset (2008-2020) (Zeng et al., 2022) were selected as the data sources for the study. The first dataset is used as the primary data source and the second dataset is used as a supplementary data source to improve the generalization capability of the model. Based on the latitude and longitude of landslide sites and landslide vector boundary files provided in the two datasets, time-series optical remote sensing images, topographic feature data, geological data, and meteorological data for the corresponding landslide sites in 2008, 2011, 2014, 2017 and 2020 were collected with the study time range of 2008-2020. Among them, optical remote sensing image data from open source Google Earth images, and the landslides on the images are manually labeled. Terrain features were selected as representative of ground relief and generated by ASTER GDEM. Geological data were selected as representative of lithology and obtained by vectorized geological maps. The meteorological data were selected as a representative of precipitation and obtained by interpolation of monitoring data from meteorological stations. Since heterogeneous data from multiple sources follow different protocols and cannot be directly input to the model for training, we perform data normalization, projection transformation, geometric correction, image cropping, and bitmap generation from different sources to obtain raster data of equal range and resolution to establish data consistency. In this study, the resolution of optical remote sensing images and other environmental factor feature data were 30m×30m, and all factors were used as continuous variables. The final formed samples consisted of optical remote sensing images, landslide-related environmental factor feature data, and labels, forming a total of 3172 samples. All the produced samples were randomly divided into training samples and test samples in a ratio of 7:3.

3.2 Evaluation Indicator

The landslide identification involved in this study can be regarded as a dichotomous classification problem. Therefore, in this paper, the mean cross-merge ratio (MIoU), F1 score, and area under the subject operating characteristic curve (AUC), which are commonly used in dichotomous classification problems, are selected as model classification accuracy evaluation metrics.

1) Mean intersection ratio (MIoU): Calculated from the intersection ratio (IoU), which is the intersection of the predicted and actual regions divided by the concatenation of the predicted and actual regions, and then the average of the intersection ratios of all categories to obtain the MIoU value. The higher the MIoU value, the better the overall classification result.

2) The area under the subject working characteristic curve (AUC): The subject working characteristic curve (ROC) and the area under the curve (AUC) have been widely used in the evaluation of the accuracy of the landslide identification model. the AUC ranges from 0 to 1, and the larger the AUC value is, the higher the correct rate and the better the accuracy of the classification model, and the worse the accuracy on the contrary.

3) F1 score: For the dichotomous problem, there are four different combinations of model final prediction results and true labels: TP, FP, TN, and FN, as shown in Table 1. F1 score

calculation relies on two basic indicators Recall and Precision, which is an indicator that better reflects the comprehensive level of the model and can be regarded as the evaluation given by the combined accuracy and completeness of the check. The higher the F1 score, the better the overall classification result. The calculation formulas are as follows.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \quad (12)$$

True value	Predicted value	
	Positive	Negative
Positive	True Positive(TP)	False Negative(FN)
Negative	False Positive(FP)	True Negative(TN)

Table 1. Combined relationship between true labels and predicted results.

3.3 Comparison Experiment

In addition to the DCNN-DBN model proposed in this paper, we selected three typical semantic segmentation models (SegNet, DeepLab V3, PSPNet) and used the same dataset for training and testing. First, a preliminary hyperparameter is determined for the network structure in the model training process, and then the trial-and-error method is used to fine-tune the hyperparameter, and the relatively optimal hyperparameter is used for landslide hazard susceptibility analysis. The final set of optimal parameters is obtained, where the middle layer activation function uses ReLU, the optimizer uses Adam, the batch size is set to 68, each batch iterates 230 times, the learning rate is 0.0001, the momentum value is 0.9, and all samples are resampled to 512×512 pixels during training.

The experimental results are shown in Figure 5. The four columns (a), (b), (c), and (d) are the extraction results corresponding to the four typical optical remote sensing images of landslides and the four network models we selected, and the landslides contained in these four images represent several major landslide scales and morphologies in the dataset. As can be seen from the areas selected by the red box, our proposed model extracts most of the landslide areas and misses only a few of them. The identification results are closest to the extent of the actual landslides, with no noise in the landslide areas and some of the closer landslides can be accurately distinguished from the boundaries. The whole landslide map identification is complete, and landslides of different scales can be identified. Compared with the extraction results of the PSPNet model and the results of the model in this paper, the overall recognition effect on visual inspection is not much different, and most of the landslides can be recognized. However, some landslides are not identified in the interior and edges, and the recognition is not as complete as the model proposed in this paper. The remaining two models were also able to identify more landslides overall, but the identification was less complete than the first two models.

To evaluate the extraction results of the four models more precisely, the extraction results were quantified using the three met-

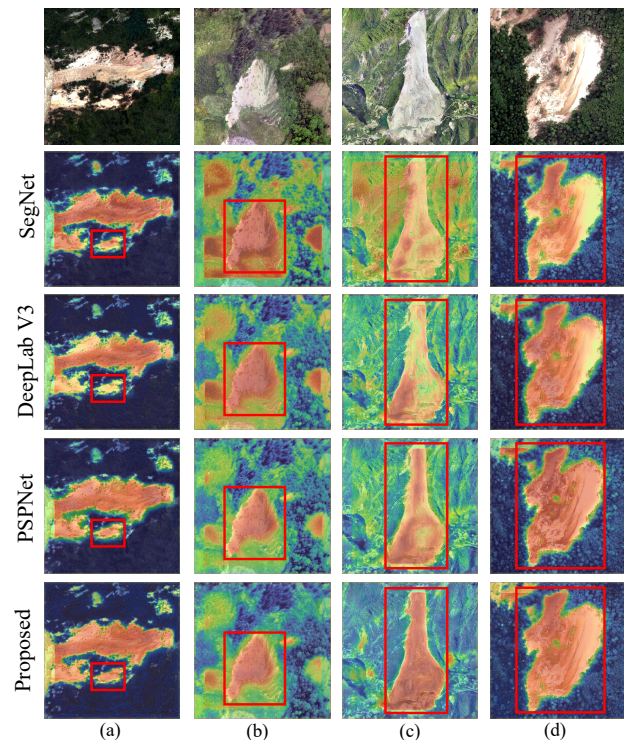


Figure 5. Comparison diagram of splitting results.

Network	Recognition accuracy	Recognition precision	
	MIoU	F1	AUC
SegNet	0.6901	0.6654	0.7139
DeepLab V3	0.7235	0.7313	0.7869
PSPNet	0.8053	0.8211	0.8137
Proposed	0.8624	0.8352	0.8976

Table 2. Landslide recognition evaluation results of four models.

rics mentioned in 4.2, and the results are shown in Table 2. The evaluation results showed that the DCNN-DBN model had an AUC value of 0.8976, an F1 score of 0.8352, and an MIoU of 0.8624. Compared with the classical network SegNet, the DCNN-DBN model improved by 0.1837, 0.1698, and 0.1723 in three metrics, AUC, F1 score, and MIoU, respectively. It can be seen that the model can better extract the features of various types of landslides, the extraction results have high marginal integrity, and the DCNN-DBN model has strong generalization ability, which provides the possibility of fast and accurate regional landslide hazard investigation.

3.4 Ablation Experiment

3.4.1 Validation of Landslide-Related Factors: To verify whether the landslide correlation factor feature data added to the model input improves the performance of the model, we set up three sets of experiments, i.e., optical remote sensing image, landslide correlation factor, and optical remote sensing image + landslide correlation factor are input to our proposed model for testing. The quantitative evaluation results of the three sample input methods are shown in Table 3 and Figure 6. It can be seen that the model using optical remote sensing image + landslide correlation factor as data input (Ours) has the best performance on all three metrics, and its AUC value and F1 score are improved by 0.1135 and 0.1063, respectively, compared with the

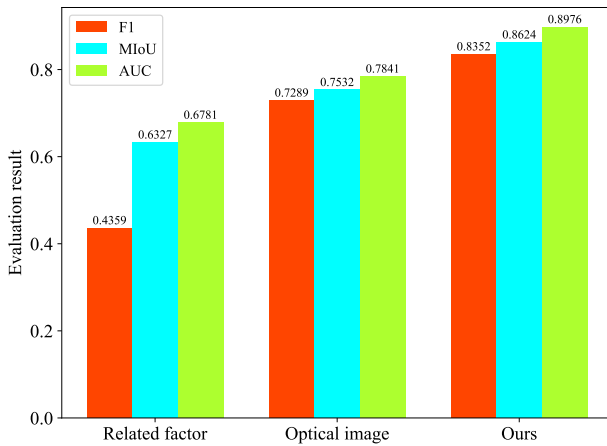


Figure 6. Evaluation results of 3 types of sample input methods.

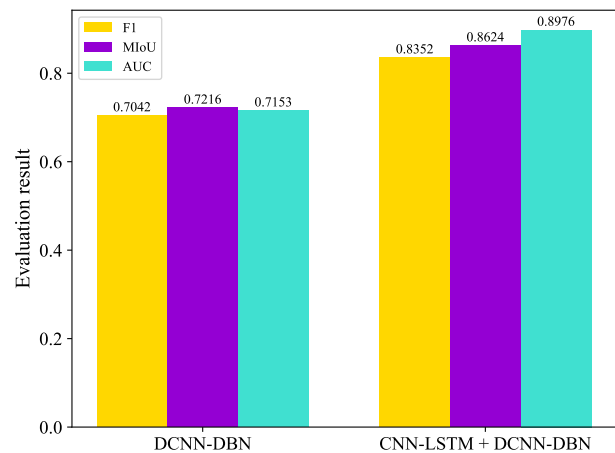


Figure 7. Evaluation results of the two groups of modules.

model using optical remote sensing image. The best recognition accuracy of our model can also be reflected from the MIOU values.

Sample input	Evaluation indicator		
	MIOU	F1	AUC
Optical image	0.7532	0.7289	0.7841
Related fator	0.6327	0.4359	0.6781
Ours	0.8624	0.8352	0.8976

Table 3. Results of landslide-related factor validation experiments.

3.4.2 Time-Phase Fusion Module Validation: To verify whether our designed CNN-LSTM-based temporal phase feature fusion module can achieve a high-dimensional representation of landslide spatio-temporal features and improve model recognition accuracy, we set up two sets of experiments. The first experiment is to input optical remote sensing images and landslide correlation factors together into the network framework based on CNN-LSTM and DCNN-DBN modules proposed in this paper. The second experiment is tested using the same data samples input to a network framework based only on the DCNN-DBN module. The experimental results were evaluated quantitatively, as shown in Table 4 and Figure 7. As can be seen in Table 4, the network framework based on CNN-LSTM and DCNN-DBN modules improves by 0.1823, 0.1310, and 0.1408 in the three metrics of AUC, F1 score, and MIOU, respectively, compared to the network framework based on DCNN-DBN modules only, with better recognition accuracy and precision. The experimental results show that our designed CNN-LSTM-based temporal phase feature fusion module is effective.

Module type	Evaluation indicator		
	MIOU	F1	AUC
DCNN-DBN	0.7216	0.7042	0.7153
CNN-LSTM + DCNN-DBN	0.8624	0.8352	0.8976

Table 4. Temporal phase fusion module validation experimental results.

4. CONCLUSION

In this paper, we propose a landslide recognition network framework based on the fusion of multi-source heterogeneous spatio-temporal data. To verify the effectiveness of the network framework proposed in this paper, we built a landslide dataset based on remote sensing images, DEM, geological data, meteorological data, and labels, and designed reasonable comparison experiments and ablation experiments to draw the following conclusions.

- 1) A network model for landslide recognition using time-series and spatially fused features from multiple sources of data is implemented. The CNN-LSTM-based temporal phase fusion module extracts and fuses the rich information in the multi-source temporal observation data of landslides, and then the DCNN-DBN spatial feature fusion module makes full use of the spatial features of each factor layer after temporal fusion to achieve efficient and accurate identification of landslides.
- 2) A comparison experiment is designed to train and test the proposed landslide recognition network framework and three other classical deep learning networks (SegNet, DeepLab V3, PSPNet) using the same dataset. The results show that the AUC value of the proposed model in this paper is 0.8976, the F1 score is 0.8352, and the mean crossover ratio (MIOU) is 0.8624, reflecting that the model can better extract the features of various types of landslides.
- 3) The quantitative evaluation results of the landslide factor feature data ablation experiment showed that the model with the combination of optical remote sensing images and landslide-related factors as data input performed the best in three indicators (MIOU, F1 score, and AUC), indicating that the input of landslide factor feature data could improve the accuracy of the model prediction.
- 4) The quantitative evaluation results of the temporal fusion module ablation experiments show that compared with the network architecture based on DCNN-DBN only, the network framework based on CNN-LSTM and DCNN-DBN modules improves the AUC value, F1 score, and MIOU value by 0.1823, 0.1310 and 0.1408, respectively, indicating that the inclusion of the temporal fusion module can improve the model's landslide recognition capability.

REFERENCES

- Chen, Y., Ming, D., Ling, X., Lv, X., Zhou, C., 2021. Landslide susceptibility mapping using feature fusion-based CPCNN-ML in Lantau Island, Hong Kong. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 3625–3639.
- Fuzhong, W., Shuaifeng, W., Li, Z., 2022. ULTRA SHORT TERM POWER PREDICTION OF PHOTOVOLTAIC POWER GENERATION BASED ON VMD-LSTM AND ERROR COMPENSATION. *Acta Energetica Solaris Sinica*, 43(8), 96.
- Hu, X., Wang, T., Pierson, T. C., Lu, Z., Kim, J., Cecere, T. H., 2016. Detecting seasonal landslide movement within the Cascade landslide complex (Washington) using time-series SAR imagery. *Remote Sensing of Environment*, 187, 49–61.
- HUANG, J., LI, X., CHEN, F., CUI, R., LI, H. et al., 2022. A deep learning recognition model for landslide terrain based on multi-source data fusion. *The Chinese Journal of Geological Hazard and Control*, 33(2), 33–41.
- Jaedicke, C., Van Den Eeckhaut, M., Nadim, F., Hervás, J., Kalsnes, B., Vangelsten, B. V., Smith, J. T., Tofani, V., Ciurcean, R., Winter, M. G. et al., 2014. Identification of landslide hazard and risk ‘hotspots’ in Europe. *Bulletin of Engineering Geology and the Environment*, 73(2), 325–339.
- Jia, X., Xiao, Y., Yang, D., Yang, Z., Wang, X., Liu, Y., 2018. Temporal-spatial feature learning of dynamic contrast enhanced-mr images via 3d convolutional neural networks. *Chinese Conference on Image and Graphics Technologies*, Springer, 380–389.
- Jiang, Y., Xu, Q., Lu, Z., Luo, H., Liao, L., Dong, X., 2021. Modelling and predicting landslide displacements and uncertainties by multiple machine-learning algorithms: Application to Baishuihe landslide in Three Gorges Reservoir, China. *Geomatics, Natural Hazards and Risk*, 12(1), 741–762.
- Keypert, R. N., Oommen, T., Martha, T. R., Sajinkumar, K., Gierke, J. S., 2018. A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images. *International journal of applied earth observation and geoinformation*, 64, 1–11.
- Li, L., Han, L., Ding, M., Cao, H., Hu, H., 2021. A deep learning semantic template matching framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 205–217.
- Li, S., Wu, L., Chen, J., Huang, R., 2020. Multiple data-driven approach for predicting landslide deformation. *Landslides*, 17(3), 709–718.
- Li, X. P., Li, Y. A., 2012. Design of GIS-based monitoring and early-warning system of landslide hazard in Diao Zhongba. *Energy Procedia*, 16, 1174–1179.
- Petley, D., Mantovani, F., Bulmer, M., Zannoni, A., 2005. The use of surface monitoring data for the interpretation of landslide movement patterns. *Geomorphology*, 66(1-4), 133–147.
- Su, X., Meng, X., Zhang, Y., Zhao, F., Yue, D., Guo, F., Zhou, Z., 2022. A dataset of catalog and characteristics of landslides along China-Pakistan Karakoram Highway. *China Scientific Data*, 7(2), 103-115.
- Wang, H., Zhang, L., Yin, K., Luo, H., Li, J., 2021. Landslide identification using machine learning. *Geoscience Frontiers*, 12(1), 351–364.
- Wu, R., Zhang, Y., Guo, C., Yang, Z., Tang, J., Su, F., 2020. Landslide susceptibility assessment in mountainous area: a case study of Sichuan–Tibet railway, China. *Environmental Earth Sciences*, 79(6), 1–16.
- Xing, H., Liu, L., 2018. Review on the new methods of landslide hazards monitoring methods in the twenty-first century. *GeoShanghai International Conference*, Springer, 134–141.
- Xu, C., Sun, Q., Yang, X., 2018. A study of the factors influencing the occurrence of landslides in the Wushan area. *Environmental Earth Sciences*, 77(11), 1–8.
- Zeng, C., Cao, Z., Su, F., Zeng, Z., Yu, C., 2022. A dataset of high-precision aerial imagery and interpretation of landslide and debris flow disaster in Sichuan and surrounding areas between 2008 and 2020. *China Scientific Data*, 7(2), 191-201.