**Engineering Letters**

# A Hybrid Deep Learning Model for Detecting and Classifying Debris Flows and Landslides in High-precision Aerial Images

Lingxuan Zhang, Xiujuan Zheng, Juan Ma, Site Mo, Fei Peng

*Abstract*—**Debris flows and landslides are geological hazards that pose a significant threat to life and property. To reduce the impact of these disasters and improve management efficiency, there is an urgent need to establish flexible and effective post-disaster management methods. High-precision aerial images offer significant advantages in this field, as their mobility and high resolution provide more detailed terrain information compared to traditional satellite images, especially in post-disaster emergency response scenarios. This paper proposes a systematic solution to detect and classify geohazard areas in high-precision aerial images. The approach uses the Attention-Pyramid U-Net (APU-Net) for aerial image segmentation, enabling accurate localization and identification of geohazard areas. For APU-Net, in the encoder part, the Atrous Spatial Pyramid Pooling (ASPP) is added in parallel to the encoder module of each layer and then connected in series with the Squeeze and Excitation (SE) module. In the decoder part, the Convolutional Block Attention Module (CBAM) is introduced after each upsampling. The APU-Net model is used to segment aerial images and extract key features. These features are then used to classify debris flow and landslide events. The experimental results show that the proposed method achieves an accuracy of 85.44% for the detection of geohazard areas and an accuracy of 76.13% for the classification of debris flows and landslides under complex terrain conditions. These results demonstrate that the proposed method provides robust technical support for disaster management by effectively detecting and classifying debris flows and landslides in high-precision aerial images.**

*Index Terms*—**Debris flow, Landslide, Detection, APU-Net, Classification.**

## I. INTRODUCTION

DEBRIS flows and landslides are common geological hazards in mountainous areas, characterized by five key features: sudden onset, tendency to occur in clusters, potential to trigger secondary disasters, seasonal patterns, and extreme destructive potential. These events typically arise from a combination of factors, such as terrain conditions, heavy rainfall, vegetation cover, climate fluctuations, and human activities. The interaction of these influences leads to varying spatial distributions of hazard-prone zones across different regions. Once triggered, such disasters often result in severe consequences, including significant loss of life and widespread damage to infrastructure [1-4].

Though both are mountain geohazards, debris flows and landslides operate under entirely different mechanical principles. Debris flows behave like violent, sediment-charged floods—often triggered by intense rainfall or sudden snowmelt—carrying over 40% solid content by volume [5,6]. They erupt abruptly, racing downhill at speeds surpassing 5 m/s, funneling massive volumes of debris (exceeding 10,000 m³) through narrow channels before spilling chaotically onto alluvial fans. Landslides, on the other hand, involve the gradual sliding of intact soil or rock masses along defined failure planes, creeping at rates of less than a meter per day. Both pose severe threats to ecosystems, infrastructure, and communities, though their destructive mechanisms diverge sharply [7].

The unpredictable behavior and often catastrophic consequences of geohazards demand highly reliable monitoring systems for both emergency response and damage assessment. This need becomes especially acute in mountainous regions, where difficult terrain and complex geology make traditional field surveys particularly challenging. Modern remote sensing approaches have emerged as practical alternatives that overcome these obstacles.

Traditional geohazard monitoring methods suffer from three key weaknesses: (1) infrequent data collection, (2) insufficient detail in spatial coverage, and (3) logistical difficulties in rough terrain. High-resolution aerial imaging addresses these shortcomings by providing centimeter-scale detail and adaptable deployment options, allowing comprehensive terrain analysis even in hard-to-reach areas. These capabilities make it especially useful for disaster response scenarios where speed and accuracy are critical.

Cutting-edge monitoring solutions combining computer vision and machine learning are transforming how we assess disasters. For instance, image-based flood monitoring systems employ real-time water level estimation through specialized image processing modules, significantly enhancing early warning capabilities [8]. Recent advancements include UAV-based aerial photogrammetry

---

Manuscript received February 8, 2025; revised July 10, 2025.

This work was supported by the fund of Technology Innovation Center for Geohazard Monitoring and Risk Early Warning, Ministry of Natural Resources (No. TICGM-2023-07) and the Geological Survey Program of China (grant number DD20211364).

Lingxuan Zhang is a postgraduate student of the College of Electrical Engineering, Sichuan University, Chengdu, 610065, China (e-mail: 1345159953@qq.com).

Xiujuan Zheng is an associate professor of the College of Electrical Engineering, Sichuan University, Chengdu, 610065, China (corresponding author, phone: 86-18613220591; e-mail: xiujuanzheng@scu.edu.cn).

Juan Ma is a manager of the China Institute of Geo-Environment Monitoring and the Technology Innovation Center for Geohazard Monitoring and Risk Early WarningMinistry of Natural Resources, Beijing, 100081, China (e-mail: majuan@mail.cgs.gov.cn).

Site Mo is an associate professor of the College of Electrical Engineering, Sichuan University, Chengdu, 610065, China (e-mail: mosite@126.com).

Fei Peng is an associate professor of the College of Electrical Engineering, Sichuan University, Chengdu, 610065, China (e-mail: pengfei@scu.edu.cn).

for infrastructure assessment [9], edge detection algorithms (Canny, Sobel) for critical structure monitoring [10], and AI-powered landslide tracking systems utilizing 4G networks [11]. Furthermore, the fusion of satellite imagery with predisposing factors has enabled predictive modeling through deep learning architectures like Deep Belief Networks [12], while edge computing solutions (e.g., K210 processors) have facilitated real-time hazard recognition [13].

Contemporary monitoring systems increasingly rely on sophisticated image classification architectures to achieve operational precision. In disaster response scenarios, these algorithms prove particularly valuable for differentiating hazard types and quantifying damage severity - capabilities that extend beyond geological emergencies to diverse industrial applications. Recent innovations demonstrate this cross-domain potential: Li's team engineered a bespoke CNN variant that simultaneously extracts and classifies features in pulmonary tissue patches, eliminating manual preprocessing through automated neural processing [14]. Long's comparative analysis of chest X-ray classification revealed data augmentation's critical role in boosting pneumonia detection accuracy across multiple CNN configurations [15]. At the intersection of environmental monitoring and machine vision, Kumar's group fused satellite telemetry with convolutional features through an innovative CNN-ANN hybrid architecture [16]. Network security applications also benefit from these advances, as shown by Wang's parallel RNN-CNN framework that deciphers encrypted traffic patterns directly from raw TLS data streams [17]. The automotive sector has seen similar progress, with Shoaib's enhanced Faster R-CNN model overcoming persistent challenges in detecting obscured objects during autonomous navigation [18]. Even waste management systems are evolving through intelligent classification, evidenced by Hasan's prototype smart receptacle integrating multi-sensor inputs with real-time material sorting algorithms [19]. These cross-domain advancements in computer vision underscore an urgent need for specialized solutions in geohazard differentiation - particularly in distinguishing morphologically similar yet mechanistically distinct events like debris flows and landslides.

This diagnostic challenge manifests acutely in mountainous terrain, where debris flow patterns and landslide morphology frequently coexist but demand divergent mitigation strategies. Current methodological approaches have coalesced around four paradigms: (i) traditional feature-classifier pipelines requiring manual geomorphometric parameterization, (ii) end-to-end CNN architectures automating feature learning, (iii) transfer learning frameworks adapting pretrained vision models, and (iv) segmentation-driven classifiers enabling pixel-level terrain parsing. Each paradigm carries distinct operational trade-offs - for instance, while CNNs excel in automated feature extraction, their performance degrades significantly when confronted with scale variance inherent to natural landslide environments. Precision demands in disaster scenarios increasingly favor segmentation-based approaches, which mitigate scale sensitivity through localized feature analysis. Early implementations achieved 70% landslide detection accuracy using SVM-classified segments in controlled forest environments [20], though struggled with complex terrain heterogeneity. Subsequent innovations introduced hybrid methodologies, exemplified by Keyport's fusion of K-means clustering with object-based image analysis [21], and Du's systematic benchmarking establishing GCN and DeepLab v3 as optimal for landslide semantic parsing (54.2% mIoU) [22]. Nevertheless, as Martha's adaptive parameter selection framework revealed [23], fundamental challenges persist in balancing segmentation granularity with computational efficiency across diverse geomorphological contexts.

Recent advancements in deep learning have solidified U-Net's position as a benchmark architecture across multiple imaging domains. Its proven efficacy spans biomedical diagnostics (demonstrated in tumor boundary delineation by Song et al. [24] and cellular structure analysis by Ibtehaz et al. [25]), environmental surveillance (as applied to deforestation tracking by Malik et al. [27]), and remote sensing applications (particularly in land-use classification as shown by Shamsolmoali et al. [26]). Despite these widespread successes, the architecture's potential remains underexplored for critical geomorphological analyses - particularly in debris flow characterization and landslide pattern recognition. U-Net's inherent advantages, including hierarchical feature learning, depth scalability without degradation risks, and precise localization capabilities, position it as an ideal candidate for these geohazard imaging challenges. Harnessing these capabilities could revolutionize our capacity to predict and manage slope instability events through computational image analysis.

Building on this technological gap, we introduce APU-Net - an Attention-Pyramid U-Net architecture specifically optimized for geohazard imagery. Our encoder enhancement integrates complementary attention mechanisms: CBAM for spatial-channel feature refinement and SE blocks for channel-wise feature recalibration. The CBAM component operates through sequential channel and spatial attention gates, dynamically emphasizing tectonically significant features while suppressing geological noise. Concurrently, the SE modules employ global context squeezing and excitation to amplify critical channel responses, effectively filtering redundant spectral information. For decoder optimization, we implement an ASPP-enhanced pyramid structure that combats context fragmentation through parallel atrous convolutions. This multi-scale approach preserves feature resolution while expanding the effective receptive field to encompass crucial geomorphological patterns. Our research makes three substantive contributions to geohazard analysis:

1. We propose a new APU-Net segmentation framework that integrates ASPP and SE modules in parallel within the encoder. This design improves the ability of the model to capture multiscale contextual features, significantly increasing accuracy in detecting geohazards under complex terrain conditions.
2. The decoder is improved by adding CBAM model after each upsampling step. By combining spatial and channel attention, the model better highlights important regions and structural boundaries, leading to more precise segmentation results.

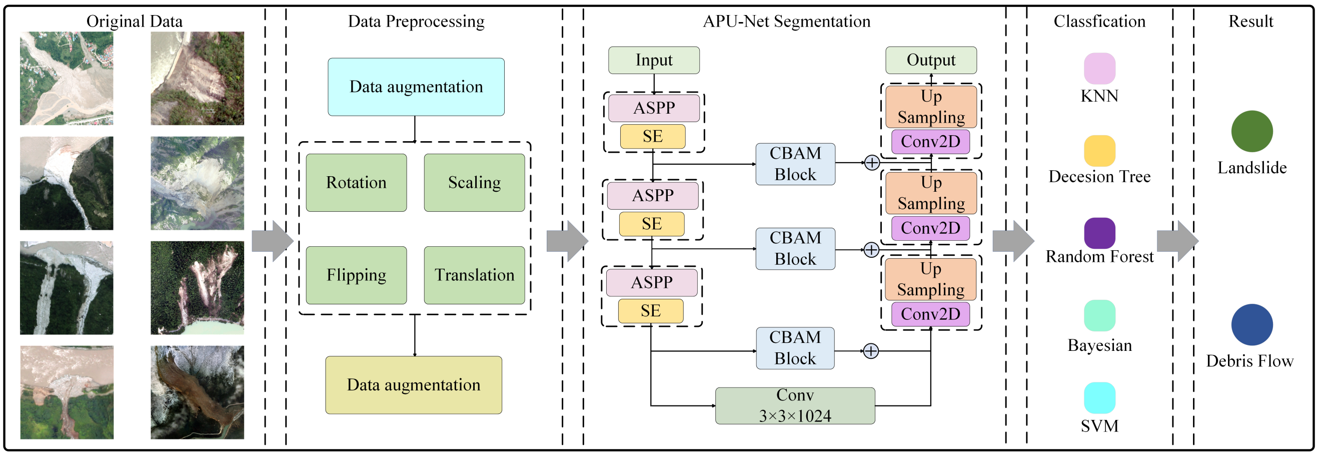

Fig. 1. Flowchart of the hybrid deep learning model-based method for debris flow and landslide detection and classification in high-precision aerial images

3. An end-to-end geohazard analysis pipeline is developed, which first segments aerial images using the proposed APU-Net, then classifies debris flow and landslide areas based on extracted features. Experiments show that this approach achieves high accuracy and robustness in both segmentation and classification, providing reliable technical support for post-disaster emergency response using high-resolution aerial imagery.

The rest of this paper is organized as follows. Section II introduces the APU-Net model. Section III describes the details of the experiments. Section IV presents the experimental results and discusses them. Finally, Section V lists the conclusions of this study.

## II. METHODOLOGY

The methodology begins with pre-processing the images through enhancement and normalization to improve their quality and ensure consistency across the dataset. This step is essential for preparing the images for accurate analysis. Following preprocessing, the APU-Net network is used for segmentation. The APU-Net model uses its advanced architecture to segment images and produce detailed segmentation masks that highlight the areas of interest. Once the segmentation masks are obtained, feature extraction is performed to derive meaningful features from the segmented images. These features capture important attributes that are relevant for further analysis. After extraction, a feature selection process is conducted to refine the dataset by identifying and retaining only the most significant features, which enhances the performance of the classification model. Finally, the selected features are utilized in classification tasks to accurately identify and categorize geohazards. This methodical approach, encompassing preprocessing, segmentation, feature extraction, and selection, ensures a robust and precise analysis, leading to effective disaster assessment and response. The overall process flowchart of the proposed method is shown in Fig 1.

### *A. Data augmentation and normalization*

To enhance the dataset for subsequent work, various data augmentation techniques were applied. These included rotation, translation, scaling, and flipping. By performing these geometric transformations on the original images, new training samples were generated, thereby improving the generalization ability and robustness of the model.

These transformations effectively expanded the dataset by creating new samples from the original images, which in turn enhanced the model's ability to generalize. By incorporating variations in orientation, position, size, and symmetry, the model became more robust in recognizing objects under diverse conditions. This augmentation process played a key role in improving the accuracy and resilience of the model when faced with real-world data.

There are several methods available for data normalization, including min-max scaling, Z-score standardization, and mean normalization. For the purposes of this study, the dataset is normalized using the Z-score standardization method. This approach is chosen due to its simplicity and ease of computation, as well as its ability to effectively normalize data regardless of the scale or presence of extremely large or small values. The Z-score standardization formula utilized is as follows:

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

where $x$ represents the data mean, $\mu$ represents the standard deviation, and $z$ represents the standardized score.

The U-Net is a convolutional neural network architecture originally designed for medical image segmentation tasks [28]. The design of U-Net is particularly suitable for tasks requiring precise localization, and it has demonstrated outstanding performance across various image segmentation tasks [29].

The APU-Net segmentation model proposed in this paper has an overall network architecture similar to the U-Net segmentation model, which mainly includes encoder, jump connection and decoder parts. On the basis of retaining the basic structure of the U-Net network, in the encoder part, CBAM is added in parallel to the encoding module of each layer, and then connected in series with the SE module. In the decoder part, the ASPP module is introduced after each upsampling to perform multi-scale perception of the feature map and improve the segmentation accuracy. The structure of APU-Net is shown in Fig 2.

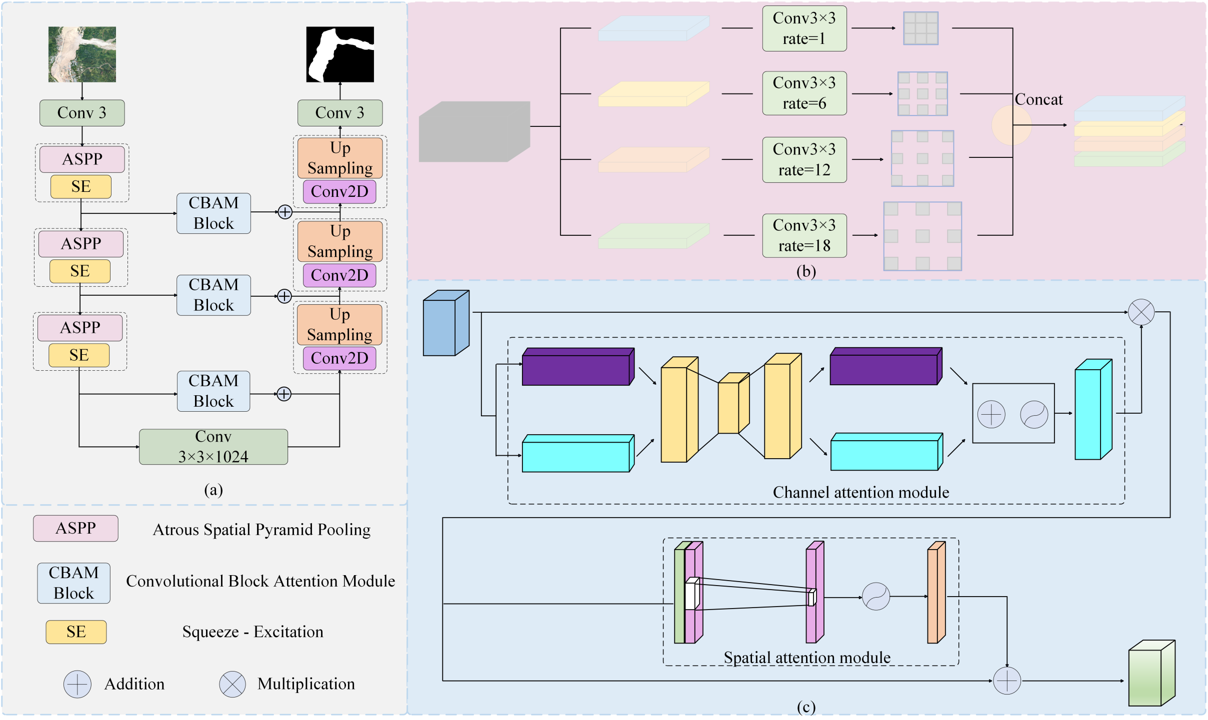

Fig. 2. (a)The Structure of Attention-Pyramid U-Net. (b)The Structure of ASPP. (c)The Structure of CBAM

### *B. Geohazard area segmentation*

*1) Atrous Spatial Pyramid Pooling:* The development of Spatial Pyramid Pooling (SPP) [30] marked a critical advancement in handling variable-resolution inputs for convolutional networks. Prior architectures constrained input dimensions through destructive resizing operations, introducing geometric distortions that compromised feature learning. SPP's breakthrough lay in its hierarchical pooling strategy, which aggregates multi-scale spatial information through progressively coarser grid partitions. This pyramid representation maintains aspect ratio fidelity while producing fixed-length descriptors, particularly beneficial for analyzing geohazard imagery where preservation of topographic relationships proves essential.

Building upon SPP's foundation, Atrous Spatial Pyramid Pooling (ASPP) [31] introduced dilated convolution kernels to expand receptive fields without sacrificing resolution. Our implementation employs three parallel dilation rates (r=6,12,18) combined with global context features, creating a multi-scale feature synthesis expressed through:

$$Y_1 = \text{Concat}\left(\text{image}(X), H_{1,1}(X), H_{6,3}(X)\right) \tag{2}$$

$$Y_2 = \text{Concat}\left(H_{12,3}(X), H_{18,3}(X)\right) \tag{3}$$

$$Y = Y_1 \oplus Y_2 \tag{4}$$

where $H_{r,n}(X)$ represents the dilated convolution operation with a dilation rate of $r$ and a kernel size of $n{\times}n$ on $X$; $\text{image}(X)$ represents the image-level features extracted from $X$ using global average pooling.

The structure of ASPP is illustrated in Fig 2(b). The input image undergoes 1×1 convolution, pooling pyramid operation (three 3×3 dilated convolutions), and dilated pooling layer operation (pooling, 1×1 convolution, and upsampling). The results are then concatenated. The pooling pyramid uses dilated convolutions with dilation rates of 6, 12, and 18 to capture features at different receptive fields, enabling multi-scale feature extraction. Incorporating ASPP into the encoder-decoder architecture enhances feature extraction, thereby efficiently capturing multi-scale information.

*2) The Convolutional Block Attention Module:* The Convolutional Block Attention Module (CBAM) [32] is a hybrid attention mechanism that combines spatial attention and channel attention mechanisms. Its structure is shown in Fig 2(d). Given an input feature map $F$ of size $C \times W \times H$, $F''$ represents the output of CBAM. The symbol $\odot$ denotes the multiplication of elements in a way that is equivocal, $cM$ represents the extraction operation of features of attention in the spatial dimension, and $sM$ indicates the attention feature extraction operation in the channel dimension.

$$F' = M_c(F) \otimes F \tag{5}$$

$$F'' = M_s(F') \otimes F' \tag{6}$$

$$F \in \mathbb{R}^{H\times W\times C},\ \ M_c \in \mathbb{R}^{1\times 1\times C},\ \ M_s \in \mathbb{R}^{H\times W\times 1} \tag{7}$$

CBAM is a hybrid attention mechanism that enhances feature representation by combining spatial attention and channel attention mechanisms. Specifically, the channel attention mechanism first weights the channel dimension of the input feature map to focus on important channels, followed by the spatial attention mechanism that weights the spatial dimension within each channel to highlight important regions in the image. Using this approach, CBAM adaptively selects crucial channels and spatial areas, thus improving the feature representation ability of the model.

*3) Squeeze-and-Excitation:* The SE module is an attention mechanism designed to enhance channel feature representation in convolutional neural networks (CNNs). It explicitly models the interdependencies between channels and dynamically adjusts the importance of each channel, thereby improving the network's feature representation capability. The SE module consists of two main stages: squeeze and excitation. The structure of SE is illustrated in Fig 2(c).

In the Squeeze stage, the SE module performs a global average pooling operation to compress spatial information from each channel into a scalar. Specifically, for a given input feature map $X \in \mathbb{R}^{H\times W\times C}$ (where $H$ is the height, $W$ is the width, and $C$ is the number of channels), global average pooling computes the average value of each channel to generate a descriptor for each channel.

In the Excitation stage, the SE module learns the inter dependencies between channels by using fully connected layers, generating attention weights for each channel. First, the compressed descriptor $z \in \mathbb{R}^{C}$ is mapped to a smaller intermediate dimension using a fully connected layer, and then it is restored to the original channel dimensions using another fully connected layer.

The SE module first compresses the spatial information, then the excitation stage learns the interdependencies between channels to generate attention weights, and finally, the recalibration step adjusts each channel's importance. This process allows the network to automatically adjust the attention given to different channels, improving its ability to respond to important features and improving the overall representation of the features [33].

### C. Geohazard classification

Feature selection is crucial to obtaining comprehensive information on debris flows and landslides. To ensure that the classification model can accurately capture relevant patterns, we have selected the following key features, as shown in the table. This process not only reduces the dimensionality of the dataset and minimizes the risk of overfitting but also eliminates redundant data, thereby improving both the model's performance and interpretability.

TABLE I
FEATURE CATEGORIES AND NAMES

| ID | Feature Category | Feature Name |
|---|---|---|
| 1 | Geometric Feature | Area |
| 2 | Geometric Feature | Length |
| 3 | Geometric Feature | Aspect Ratio |
| 4 | Texture Feature | Contrast |
| 5 | Texture Feature | Dissimilarity |
| 6 | Texture Feature | Homogeneity |
| 7 | Texture Feature | Energy |
| 8 | Color Feature | Color Histogram |
| 9 | Color Feature | Gradient Color Feature Mean |
| 10 | Color Feature | Gradient Color Feature Standard Deviation |
| 11 | Grayscale Feature | Grayscale Mean |
| 12 | Grayscale Feature | Grayscale Standard Deviation |

Our feature engineering strategy employs an iterative refinement process combining backward elimination with SHAP value analysis. This dual approach maintains methodological transparency while establishing an empirical basis for feature importance quantification. The SHAP framework's game-theoretic foundation proves particularly valuable for disentangling complex feature interactions in geohazard prediction tasks, where multivariate dependencies often obscure individual parameter contributions. Through successive iteration cycles, we systematically evaluate each feature's marginal utility under controlled ablation conditions, ensuring the final feature set captures essential discriminative patterns without redundancy.

In this study, we used several machine learning models, including K-nearest neighbors (KNN), support vector machine (SVM), random forest, naive Bayes and decision tree classifiers, to classify geohazards using the selected features. These models were chosen for their distinct advantages and complementary strengths in handling different aspects of classification tasks. Using these features, we were able to leverage the unique capabilities of each model to enhance classification performance and gain deeper insights into geohazard characteristics.

The decision to select these models was driven by their proven performance in similar classification tasks and their ability to handle diverse data characteristics. Naive Bayes, with its foundation in probability theory, is efficient for large datasets and performs well even with relatively small amounts of data, making it suitable for situations with limited training samples. The decision tree model, on the other hand, is highly interpretable, allowing for a clear understanding of the decision-making process, which is valuable in identifying the key factors contributing to geohazard classifications. By using a combination of these models, we aimed to leverage their strengths to achieve accurate and reliable geohazard classification results.

## III. EXPERIMENTS

### A. Data describtion

The dataset employed in this study builds upon the work of Zeng et al. [34], who assembled an extensive collection of data on landslide and debris flow disasters within Sichuan and its neighboring areas. This dataset incorporates a wide array of data sources, interpretation techniques, and comprehensive documentation of disaster events, with a particular emphasis on regions impacted by significant seismic activity and coastal areas. It includes 107 high-resolution images that capture typical landslide and debris flow disasters, each meticulously annotated to facilitate the analysis and classification of these geohazards.

The dataset is derived from aerial orthophotography data collected from both unmanned aerial vehicles (UAVs) and manned aircraft, focusing on landslide and debris flow disasters that have occurred in Sichuan province since 2008. The data was meticulously organized, standardized, and then subjected to visual interpretation and annotation using ArcGIS software. This comprehensive dataset is invaluable for advancing research in geohazard classification and management, as it provides high-resolution imagery and detailed annotations that enhance the accuracy and effectiveness of disaster analysis, mitigation strategies, and response planning.

Due to the limited size of the original dataset, image augmentation techniques were employed to prevent suboptimal model training results caused by insufficient data. These techniques included rotation, scaling, and flipping, which diversified the original images and generated more training samples. Through these augmentation methods, the original 107 images were expanded to 500 images, effectively increasing the dataset size.

### B. Evaluation metrics

The Dice coefficient, also known as the Sørensen–Dice coefficient, is a statistical tool used to measure the similarity between two samples. In the context of image segmentation, it is employed to compare the predicted segmentation results with the actual data. The value of the Dice coefficient ranges from 0 to 1, with 1 indicating a perfect segmentation. The calculation formula is:

$$\text{DiceCoefficient} = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{8}$$

where $X$ represents the predicted segmentation results, while $Y$ denotes the actual ground truth data.

The Intersection over Union (IoU) is another metric for assessing segmentation accuracy. It calculates the ratio of the overlapping area to the combined area of the predicted and actual segmentations. The IoU value ranges from 0 to 1, with 1 indicating a perfect segmentation. The formula for IoU is:

$$\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|} \tag{9}$$

The definitions of $X$ and $Y$ are the same as those used for the Dice coefficient.

Recall: In image segmentation, this typically refers to the model's ability to correctly identify positive classes. Higher recall indicates fewer missed positive instances. The formula for recall is:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

where $TP$ represents True Positives and $FN$ represents False Negatives.

Precision: This metric evaluates the model's ability to correctly identify negative cases. A higher precision indicates a lower rate of misclassifying negative cases as positive. The calculation formula is as follows:

$$Precision = \frac{TN}{TN + FP} \tag{11}$$

where $TN$ represents True Negatives and $FP$ represents False Positives.

The False Rejection Rate (FRR) is a metric used to assess the accuracy of a model, especially in classification and recognition tasks. It measures the proportion of instances where the model incorrectly rejects a valid positive case. In the context of image segmentation or classification, it refers to the rate at which positive instances are misclassified as negatives. The formula for calculating the False Rejection Rate is:

$$FRR = \frac{FN}{FN + TP} \tag{12}$$

where $FN$ represents False Negatives and $TP$ represents True Positives.

### C. Implementation details

In the U-Net architecture, the number of convolutional layers, the size of convolutional kernels, and the choice of activation functions are critical hyperparameters in model design. Conducting ablation experiments, such as varying convolutional kernel sizes, activation functions, and the number of convolutional layers, facilitates a deeper understanding of their impact on model performance. Through ablation experiments, it is possible to systematically evaluate the contribution of each component to the model's performance, thereby providing robust guidance for model design and optimization.

A smaller number of convolutional layers may result in the model's inability to capture sufficient hierarchical features, yet it offers faster training speed, fewer parameters, and is suitable for simple tasks or small datasets. Conversely, increasing the number of convolutional layers enables the capture of more complex hierarchical features but demands more computational resources, making it suitable for complex tasks or large datasets. The size of convolutional kernels influences the local or global capture of feature information, where smaller kernels preserve spatial resolution and larger kernels provide a broader contextual understanding. Activation functions like ReLU facilitate rapid training and robustness against noise, whereas LeakyReLU can mitigate the issue of dying ReLU.

By comprehensively considering these parameters, it becomes possible to balance the model's performance and efficiency, thus guiding model design and optimization effectively. To attain the optimal combination of hyperparameters, the network structure of the APU-Net is configured as follows:

TABLE II
EXPERIMENTAL PARAMETER SETTINGS

| ID | Number of Convolutional Layers | Convolutional Kernel Size | Activation Function |
|---|---|---|---|
| 1 | 2 | 3×3 | ReLU |
| 2 | 3 | 3×3 | ReLU |
| 3 | 4 | 3×3 | ReLU |
| 4 | 2 | 5×5 | ReLU |
| 5 | 3 | 5×5 | ReLU |
| 6 | 4 | 5×5 | ReLU |
| 7 | 2 | 3×3 | LeakyReLU |
| 8 | 3 | 3×3 | LeakyReLU |
| 9 | 4 | 3×3 | LeakyReLU |
| 10 | 2 | 5×5 | LeakyReLU |
| 11 | 3 | 5×5 | LeakyReLU |
| 12 | 4 | 5×5 | LeakyReLU |

Following the ablation experiments with APU-Net, the next step involves selecting the best-performing model based on its evaluation metrics. This model will be used for feature extraction, allowing us to capture the most relevant features from the data. Once features are extracted, we will proceed with a systematic feature selection process to identify the most influential variables that contribute to model performance. This step ensures that only the most pertinent features are retained, enhancing the efficiency and accuracy of subsequent classification tasks. Finally, we will apply various classification algorithms to the selected features to classify geohazards effectively. This comprehensive approach

Fig. 3. Experimental results of image segmentation using different models

aims to refine the model's performance and achieve robust and reliable classification outcomes.

## IV. RESULTS

### A. *Results of geohazard area segmentaion*

The Table III presents the results of the geohazard area segmentation ablation experiments, utilizing the original images and their masks as inputs:

TABLE III
SEGMENTATION EXPERIMENTAL RESULTS

| ID | Acc | Dice | IoU | R | P |
|---|---|---|---|---|---|
| 1 | 0.7732 | 0.5733 | 0.4556 | 0.5248 | 0.9194 |
| 2 | 0.7864 | 0.5478 | 0.4473 | 0.4916 | 0.9347 |
| 3 | 0.7913 | 0.6244 | 0.4937 | 0.5680 | 0.9296 |
| 4 | **0.8105** | **0.6568** | **0.5277** | **0.5787** | **0.9493** |
| 5 | 0.8096 | 0.6192 | 0.5011 | 0.5518 | 0.9416 |
| 6 | 0.8099 | 0.6189 | 0.5038 | 0.5482 | 0.9436 |
| 7 | 0.7908 | 0.5801 | 0.4510 | 0.5053 | 0.9355 |
| 8 | 0.7919 | 0.6005 | 0.4781 | 0.5266 | 0.9366 |
| 9 | 0.7752 | 0.6121 | 0.4840 | 0.5756 | 0.9082 |
| 10 | 0.8062 | 0.6371 | 0.5009 | 0.5481 | 0.9539 |
| 11 | 0.8034 | 0.6255 | 0.5102 | 0.5516 | 0.9440 |
| 12 | 0.7898 | 0.6042 | 0.4881 | 0.5716 | 0.9128 |

Based on the table, it can be observed that when increasing the depth from 2 layers to 4 layers, almost all performance metrics of the model improve. This indicate that increasing the number of layers helps the model learn more complex features. However, as the depth of the network increases, the complexity of the model also increases, leading to an exponential increase in the number of relevant parameters within the model. This approach incurs significant consumption of computer memory and training time. Considering these factors collectively, a depth of 4 layers is deemed appropriate.

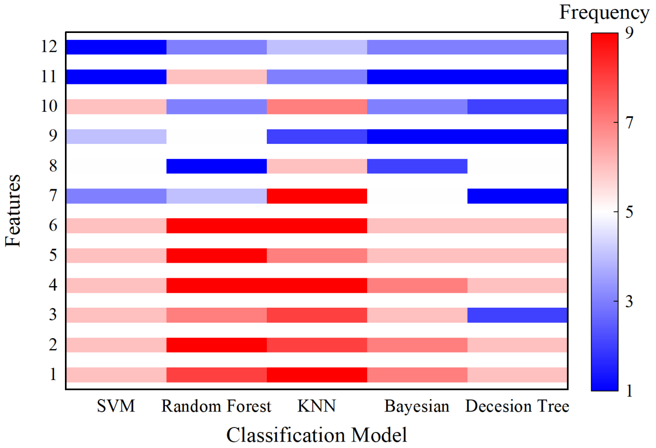

Fig. 4. The frequency of occurrence of features in each classification model

Models using a 5×5 convolutional kernel (experiments 4, 5, 6) outperform those using a 3×3 convolutional kernel in terms of Dice coefficient and IoU. In experiments 7 to 12, models utilizing LeakyReLU instead of ReLU did not show significant performance improvements or declines across most metrics. This suggests that variation in activation functions is not the primary determinant of performance.

Through these ablative experiments, it can be concluded that increasing the appropriate number of convolutional layers and utilizing larger convolutional kernel sizes can enhance the model's performance, while the variation in activation functions has minimal impact on the results.

We used the APU-Net model with a depth of 4, a 5×5 convolution kernel, and ReLU as the activation function, comparing it with several traditional segmentation methods. The comparison is shown in Fig 3.

The comparative evaluation demonstrates APU-Net's superior segmentation capability, achieving the highest accuracy at 0.8544 and Dice coefficient at 0.7768. While CBAM-SE-U-Net shows competitive precision at 0.8806 and

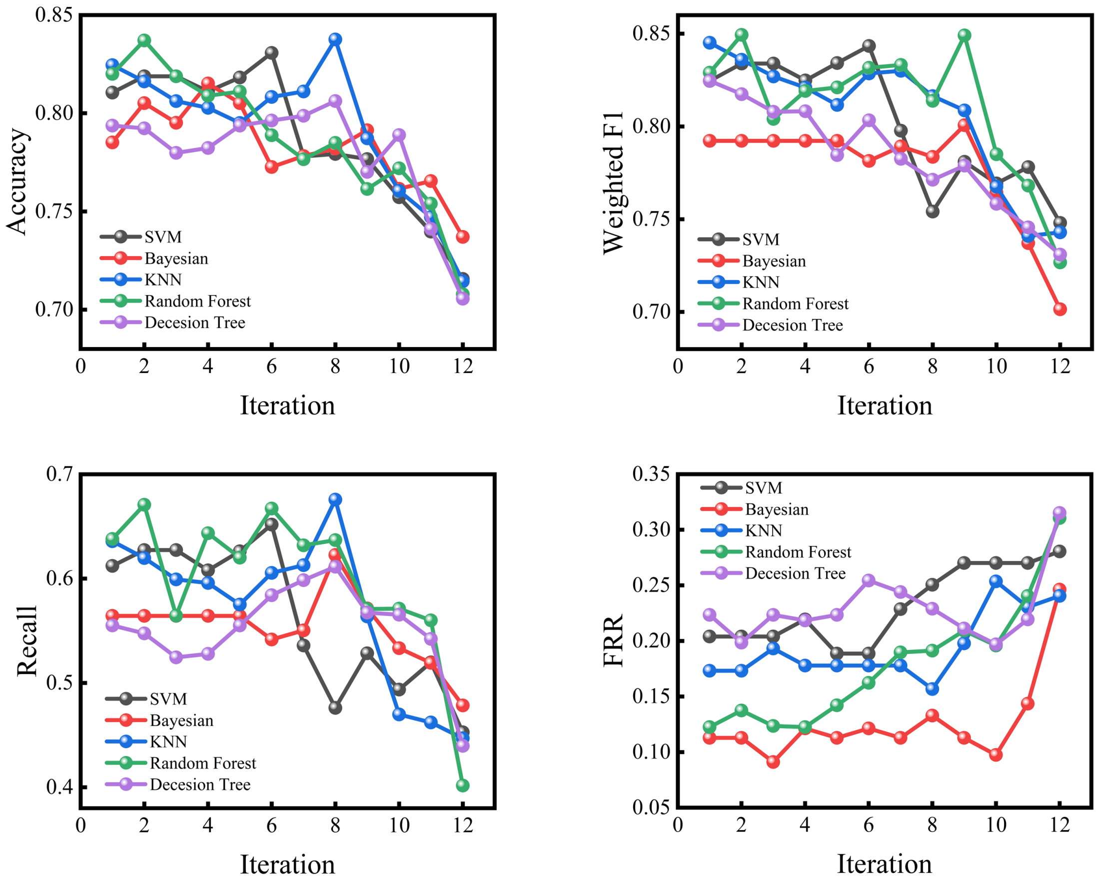

Fig. 5. Experimental results of image classification using different models

TABLE IV
SEGMENTATION PERFORMANCE COMPARISON OF DIFFERENT MODELS

| Model | Acc | Dice | IoU | R | P |
|---|---|---|---|---|---|
| Deeplabv3 | 0.8192 | 0.6641 | 0.5184 | 0.5654 | 0.8312 |
| FCN | 0.7879 | 0.5625 | 0.4738 | 0.5518 | 0.6856 |
| U-Net | 0.8235 | 0.7110 | 0.5502 | 0.6361 | 0.8408 |
| CBAM-SE-U-Net | 0.8503 | 0.7531 | 0.5818 | 0.6798 | **0.8806** |
| ASPP-U-Net | 0.8469 | 0.7501 | **0.6067** | 0.6868 | 0.8605 |
| APU-Net | **0.8544** | **0.7768** | 0.6022 | **0.7155** | 0.8721 |

ASPP-U-Net leads in IoU at 0.6067, APU-Net maintains the most balanced performance profile. The 7.15% recall advantage over baseline U-Net particularly highlights its improved detection of subtle geohazard features. The visual analysis in Figure 3 confirms these quantitative findings, with APU-Net producing segmentation boundaries that most accurately preserve the complex topographical signatures characteristic of landslide margins, while other models exhibit oversmoothing or fragmentation artifacts. This performance advantage stems from APU-Net's optimized architecture combining appropriate depth, receptive field size, and efficient feature aggregation.

### *B. Results of geohazard classification*

We conducted classification experiments using Decision Tree, Random Forest, KNN, Bayesian, and SVM. The SHAP values from 12 iterations were averaged, and features with values above the average were statistically analyzed, as shown in Fig 4. The results indicate that Features 1, 2, 4, 5, and 6 exhibited high occurrence frequencies across all classification models, as indicated by the deep red coloration, suggesting their high importance in the feature selection process. Feature 3 followed with a slightly lower frequency. The remaining features demonstrated the lowest frequencies, appearing predominantly in dark blue across the models, indicating their limited contribution to the classification task.

These results are consistent with the basic geomorphological relationship. The advantages of texture contrast reflect the significant alternating light and dark characteristics caused by surface structure rupture in debris flow and landslide images. This high-contrast texture pattern is a typical characterization of geological disaster bodies. The selection of geometric features reveals the important role of morphological features such as aspect ratio and area, which is in line with the special geometric morphological characteristics generated during debris flow and landslide movements. The low significance of grayscale features indicates that pure brightness information has limited discrimination power in geological disaster recognition.

In each iteration, we conduct classification experiments. Specifically, after removing one feature, the remaining features are used as input to the classification. The results classification is shown in Fig 5.

We selected the feature subsets from the eighth and ninth iterations as optimal candidates based on their performance in the sequential backward feature selection process guided by SHAP values. These subsets demonstrated a strong balance between model performance and feature simplicity in the classification of debris flow and landslide events.

From the figure, it can be seen that the model performance did not decrease significantly from the first to the seventh iterations. For instance, the weighted F1 score of the Random Forest classifier dropped sharply from 0.8493 in the second iteration to 0.8040 in the third, suggesting that redundant features introduced noise and weakened generalization. Similarly, the recall rate of KNN increased from 0.63577 in the first iteration to a peak of 0.6757 in the eighth, further indicating that earlier feature sets likely contained interfering or irrelevant variables.

Although the sixth and seventh iterations showed relatively high average accuracy, their recall and F1 score performance was not good. In comparison, the eighth and ninth iterations achieved good model performance while reducing the number of features: the eighth iteration enabled KNN to achieve optimal accuracy (0.8375), recall (0.6757), and FR rate (0.1566); while the ninth iteration achieved the highest weighted F1 score for Random Forest (0.84908), and the Bayesian classifier also reached its highest F1 score (0.8006).

From the tenth to twelfth iterations, model performance declined significantly as features were reduced. When the number of features decreased to the level of the tenth iteration, all core metrics deteriorated substantially. For example, the weighted F1 score of Random Forest dropped from 0.8490 in the ninth iteration to 0.7850, KNN's recall decreased from 0.6757 in the eighth iteration to 0.4697, and the average FR rate exceeded 0.2783 by the twelfth iteration. These declines reflect the loss of critical discriminatory features.

It should be noted that between the ninth and tenth iterations, the average weighted F1 score across all models decreased sharply from 0.8036 to 0.7689, a decrease of 4.3%. This nonlinear decrease indicates the existence of a critical threshold in feature selection. The features of the eighth and ninth iterations lie precisely at the edge of this threshold, achieving maximum performance enhancement.

To further analyze the ROC curves of the eighth and ninth iterations, we focus on the AUC values as a key indicator of model discrimination. In Iteration 8, Random Forest's AUC value is the highest at 0.914, indicating their strong ability to distinguish between debris flow and landslide events. KNN also performs well with an AUC of 0.898, while SVM and Decision Tree have slightly lower but still decent AUCs of 0.872 and 0.870 respectively.

In Iteration 9, Random Forest's AUC value remains the highest at 0.922, solidifying its position as a top performer. Naive Bayes maintains a good level with an AUC of 0.905, and KNN has a slight decrease to 0.879. SVM and Decision Tree see minor drops in their AUCs to 0.870 and 0.874 respectively.

By comparing the two iterations, we observe that the Random Forest model consistently demonstrates strong performance, with its AUC increasing from Iteration 8 to 9. For Naive Bayes, the AUC remains relatively stable, indicating robustness across these iterations.

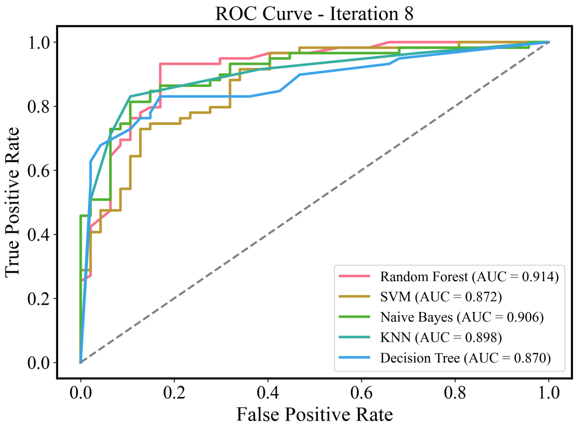

Fig. 6. ROC Curves of Models in the 8th Iteration

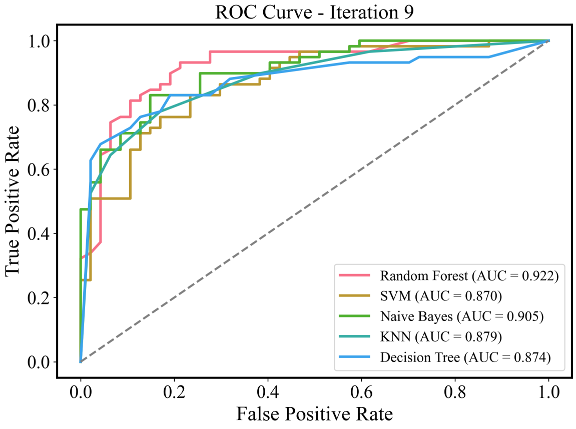

Fig. 7. ROC Curves of Models in the 9th Iteration

A curve closer to the top-left corner of the plot indicates better performance. In both iterations, Random Forest's position is closer to the top-left corner, indicating its better performance. From the heatmap, it can be seen that in the Random Forest model iteration experiments, Geometric Feature and Texture Feature appear more frequently, while Color Feature and Grayscale Feature appear less frequently. The features of the ninth iteration are length, aspect ratio, homogeneity, and energy, which is also consistent with the pattern shown in the heatmap. This indicates that Geometric Feature and Texture Feature play a more critical role in distinguishing debris flows from landslides, providing valuable guidance for feature engineering in future geohazard classification tasks.

This study systematically identifies critical patterns in machine learning-based geological hazard image recognition through feature selection and model comparisons. The significant advantages shown by random forests show that they have both characteristic robustness and classification stability. This ability to balance feature streamlining and model effectiveness makes it a preferred model for intelligent identification of geological disasters. This study provides an important basis for optimizing the characteristic engineering of the geological disaster monitoring system. In the future, the combination of deep learning and SHAP interpretation methods will be further explored to improve the accuracy of disaster recognition in complex geomorphological scenarios.

## V. CONCLUSION

In conclusion, this study addresses the critical need for effective post-disaster management of debris flow and landslide, two of the most unpredictable and destructive geohazards in mountainous regions. Given the challenges posed by their uneven distribution and complex characteristics, this research proposes a systematic approach that leverages APU-Net for precise segmentation of high-resolution aerial images. The study demonstrates that this model can accurately identify and localize geohazard areas, achieving a segmentation accuracy of 85.44% and a classification accuracy of 76.13% in complex terrain conditions. These results validate the robustness and reliability of the proposed method and demonstrate its practical application in the field of disaster management. By effectively segmenting and classifying debris flow and landslide images, this approach provides essential technical support for enhancing the efficiency and accuracy of geohazard monitoring and response efforts, ultimately contributing to more effective disaster risk mitigation strategies.

## REFERENCES

[1] F. Yang, X. Men, Y. Liu, H. Mao, Y. Wang, L. Wang, X. Zhou, C. Niu, X. Xie, "Estimation of Landslide and Mudslide Susceptibility with Multi-Modal Remote Sensing Data and Semantics: The Case of Yunnan Mountain Area," *Land*, vol. 12, no. 10, pp. 1949, 2023.

[2] R. J. Blong, "A numerical classification of selected landslides of the debris slide-avalanche-flow type," *Engineering Geology*, vol. 7, no. 2, pp. 99-114, 1973.

[3] X. Liu, C. Yu, P. Shi, W. Fang, "Debris flow and landslide hazard mapping and risk analysis in China," *Frontiers of Earth Science*, vol. 6, pp. 306-313, 2012.

[4] D. W. Park, N. V. Nikhil, S. R. Lee, "Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event," *Natural Hazards and Earth System Sciences*, vol. 13, no. 11, pp. 2833-2849, 2013.

[5] G. G. Chevalier, V. Medina, M. Hürlimann, A. Bateman, "Debris-flow susceptibility analysis using fluvio-morphological parameters and data mining: application to the Central-Eastern Pyrenees," *Natural Hazards*, vol. 67, pp. 213-238, 2013.

[6] W. Xu, W. Yu, S. Jing, Z. Wang, G. Zhang, J. Huang, "Debris flow prediction models based on environmental factors and susceptible subarea classification in Sichuan, China," *Natural Hazards*, vol. 67, pp. 869-878, 2013.

[7] Z. Y. Wang, J. H. W. Lee, C. S. Melching, "Debris flows and landslides," *River Dynamics and Integrated River Management*, pp. 193-264, 2015.

[8] C. H. Tseng, L. C. Chen, J. H. Wu, F. P. Lin, R. K. Sheu, "An automated image dehazing method for flood detection to improve flood alert monitoring system," *Journal of the National Science Foundation of Sri Lanka*, vol. 46, no. 3, pp. 329-339, 2018.

[9] J. Esteban, R. E. McRoberts, A. Fernández-Landa, J. L. Tomé, E. Nsset, "Estimating forest volume and biomass and their changes using random forests and remotely sensed data," *Remote Sensing*, vol. 11, no. 16, pp. 1944, 2019.

[10] R. Costache, Q. B. Pham, E. Sharifi, "Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques," *Remote Sensing*, vol. 12, no. 1, pp. 106, 2019.

[11] Q. Meng, C. Wang, "Application of Video Recognition Technology in Landslide Monitoring System," *MATEC Web of Conferences*, vol. 153, pp. 07001, 2018.

[12] L. Long, F. He, H. Liu, "The use of remote sensing satellite using deep learning in emergency monitoring of high-level landslides disaster in Jinsha River," *The Journal of Supercomputing*, vol. 77, no. 8, pp. 8728-8744, 2021.

[13] B. Wei, Y. Zhou, J. Huang, J. Zhang, X. Wei, "Machine Vision Monitoring of Mountain Flood Disaster and Landslide Warning Based on K210," *International Journal of New Developments in Engineering and Society*, vol. 8, no. 1, 2024.

[14] Q. Li, W. Cai, X. Wang, Y. Zhou, D. Feng, M. Chen, "Medical image classification with convolutional neural network," *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 844-848, 2014.

[15] S. S. Yadav, S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, pp. 1-18, 2019.

[16] P. Prashanth Kumar, V. Sanjana, B. Saisri, R. Kirthana, "Deep Learning for Satellite Picture Classification," *Lampyrid: The Journal of Bioluminescent Beetle Research*, vol. 13, pp. 283-289, 2023.

[17] X. Wang, S. Chen, J. Su, "App-net: A hybrid neural network for encrypted mobile traffic classification," *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops*, pp. 424-429, 2020.

[18] M. Shoaib, T. Hussain, B. Shah, I. Ullah, S. M. Shah, F. Ali, S. H. Park, "Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease," *Frontiers in Plant Science*, vol. 13, pp. 1031748, 2022.

[19] M. K. Hasan, M. A. Khan, G. F. Issa, A. Atta, A. S. Akram, M. Hassan, "Smart waste management and classification system for smart cities using deep learning," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, pp. 1-7, 2022.

[20] M. Van Den Eeckhaut, N. Kerle, J. Poesen, J. Hervás, "Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data," *Geomorphology*, vol. 173, pp. 30-42, 2012.

[21] R. N. Keyport, T. Oommen, T. R. Martha, K. S. Sajinkumar, J. S. Gierke, "A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 64, pp. 1-11, 2018.

[22] B. Du, Z. Zhao, X. Hu, G. Wu, L. Han, L. Sun, Q. Gao, "Landslide susceptibility prediction based on image semantic segmentation," *Computers & Geosciences*, vol. 155, pp. 104860, 2021.

[23] T. R. Martha, N. Kerle, C. J. Van Westen, V. Jetten, K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4928-4943, 2011.

[24] H. Song, Y. Wang, S. Zeng, X. Guo, Z. Li, "OAU-net: Outlined Attention U-net for biomedical image segmentation," *Biomedical Signal Processing and Control*, vol. 79, pp. 104038, 2023.

[25] N. Ibtehaz, M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74-87, 2020.

[26] P. Shamsolmoali, M. Zareapoor, R. Wang, H. Zhou, J. Yang, "A novel deep structure U-Net for sea-land segmentation in remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3219-3232, 2019.

[27] K. Malik, C. Robertson, D. Braun, C. Greig, "U-Net convolutional neural network models for detecting and quantifying placer mining disturbances at watershed scales," *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, pp. 102510, 2021.

[28] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, pp. 234-241, 2015.

[29] Y. Zhang, Y. Wang, Z. Zhang, "Retinal Vessel Segmentation Algorithm Based on U-NET Convolutional Neural Network," *Engineering Letters*, vol. 31, no.4, pp. 1837-1846, 2023.

[30] K. He, X. Zhang, S. Ren, J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.

[31] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2017.

[32] S. Woo, J. Park, J. Y. Lee, "CBAM: Convolutional block attention module," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3-19, 2018.

[33] S. Ge, D. Liu, X. Shi, X. Zhao, X. Wang, J. Fan, "Semantic Segmentation of Remote Sensing Images Based on Filtered Hybrid Attention Mechanisms," *Engineering Letters*, vol. 33, no. 1, pp. 80-89, 2025.

[34] C. Zeng, Z. Cao, F. Su, Z. Zeng, C. Yu, "High-precision aerial imagery and interpretation dataset of landslide and debris flow disaster in Sichuan and surrounding areas," *China Sci. Data*, vol. 7, no. 2, pp. 195-205, 2022.