



ARTICLE

# Attention Driven YOLOv5 Network for Enhanced Landslide Detection Using Satellite Imagery of Complex Terrain

Naveen Chandra<sup>1</sup>, Himadri Vaidya<sup>2,3</sup>, Suraj Sawant<sup>4</sup>, Shilpa Gite<sup>5,6</sup> and Biswajeet Pradhan<sup>7,\*</sup>

<sup>1</sup>Wadia Institute of Himalayan Geology, Dehradun, 248001, India

<sup>2</sup>Graphic Era Hill University, Dehradun, 248001, India

<sup>3</sup>Graphic Era Deemed University, Dehradun, 248001, India

<sup>4</sup>COEP Technological University, Pune, 411005, India

<sup>5</sup>Artificial Intelligence and Machine Learning Department, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, 412115, India

<sup>6</sup>Symbiosis Centre of Applied AI (SCAAI), Symbiosis International (Deemed) University, Pune, 412115, India

<sup>7</sup>Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

\*Corresponding Author: Biswajeet Pradhan. Email: biswajeet.pradhan@uts.edu.au

Received: 14 February 2025; Accepted: 27 May 2025; Published: 30 June 2025

**ABSTRACT:** Landslide hazard detection is a prevalent problem in remote sensing studies, particularly with the technological advancement of computer vision. With the continuous and exceptional growth of the computational environment, the manual and partially automated procedure of landslide detection from remotely sensed images has shifted toward automatic methods with deep learning. Furthermore, attention models, driven by human visual procedures, have become vital in natural hazard-related studies. Hence, this paper proposes an enhanced YOLOv5 (You Only Look Once version 5) network for improved satellite-based landslide detection, embedded with two popular attention modules: CBAM (Convolutional Block Attention Module) and ECA (Efficient Channel Attention). These attention mechanisms are incorporated into the backbone and neck of the YOLOv5 architecture, distinctly, and evaluated across three YOLOv5 variants: nano (n), small (s), and medium (m). The experiments use open-source satellite images from three distinct regions with complex terrain. The standard metrics, including F-score, precision, recall, and mean average precision (mAP), are computed for quantitative assessment. The YOLOv5n + CBAM demonstrates the most optimal results with an F-score of 77.2%, confirming its effectiveness. The suggested attention-driven architecture augments detection accuracy, supporting post-landslide event assessment and recovery.

**KEYWORDS:** Attention mechanism; convolutional neural networks; landslides; remote sensing images; YOLOv5

## 1 Introduction

Landslides are critical geo-hazards driven by climatic, tectonic, and human-induced activities, causing loss of life, damage to infrastructure and property, and economic disruption [1]. Given current weather patterns, urban expansion, and population growth, more landslides are anticipated in mountainous regions worldwide [2]. Therefore, effective mitigation and associated risk reduction are highly imperative for humanity. Monitoring, prediction/forecasting, and localization/detection are fundamental components in managing landslide risk [1,3,4]. Monitoring involves observing displacement or deformation in an area over an extended period, forecasting refers to predicting the future based on long-term past information, and



detection involves extracting information about the occurrence of a landslide [1]. Continuous monitoring of landslides is vital for their prediction, whereas detection leads to the identification of critical parameters, for example, site/location, which are necessary to minimize their cascading effects, aiding first responders during the post-event recovery phase.

The ability to detect landslides quickly after they occur is crucial for initiating immediate response measures. Rapid detection allows emergency responders to deploy quickly rescue teams, medical aid, and supplies to affected areas. It supports the prompt evacuation of people from hazardous zones to prevent further casualties. It aids in performing preliminary assessments of the extent of damage to infrastructure and the environment, which is essential for planning effective recovery operations. Accuracy in landslide detection is vital to ensure that the information used for decision-making is reliable. Accurate detection involves identifying true landslide events by minimizing false positives (incorrectly identified landslides) and false negatives (missed landslides) to provide a clear and precise mapping of affected areas. Accurate detection is crucial for detailed mapping, indicating the precise locations and extent of landslides, which are critical for targeted interventions. Furthermore, timeliness refers to the promptness with which data is gathered, processed, and made available to decision-makers and emergency responders. Timely data is essential as it supports immediate response efforts, reducing the time between the occurrence of a landslide and the initiation of rescue and recovery operations. It also provides real-time situational awareness, enabling responders to understand the current state of affected areas and make informed decisions. In addition, reliability pertains to the data's accuracy and consistency. Decision-makers rely on reliable and dependable data to formulate effective response strategies and allocate resources efficiently. It aids in comprehensive risk assessment, helping to identify high-risk areas and prioritize interventions. Reliable data also supports long-term planning and mitigation efforts, contributing to the development of resilient infrastructure and communities that are better prepared for future landslide events.

Moreover, detecting landslide information is important for post-event actions and inventory preparation. In addition, landslide event detection supports all the later stages of the landslide risk cycle, such as susceptibility assessment, hazard, vulnerability, and risk assessment, which are crucial for sustainable planning. Therefore, fast and accurate detection of landslide events has emerged as a recent trend in the landslide domain. The traditional landslide extraction approach was based on *in-situ* visits [5,6]. These methods were time-consuming, laborious, and less efficient in emergency conditions [6]. With the progress in remote sensing techniques, high-resolution imagery has been widely used in the studies of landslide hazard analysis [1]. Currently, four types of techniques are used for landslide detection employing remote sensing data [5]: visual/manual interpretation and analysis, pixel-based, object-based/slope unit, and artificial intelligence (AI).

Visual interpretation methods are based on the knowledge of experts, nevertheless, they require a lot of time, and struggle to meet the standards necessary for fast recovery actions [6]. Pixel-based methods overcome the limitation of visual methods by using a binary algorithm [5] to recognize the correct class or category (landslide and background) of the pixel in an image. If the objects in an image possess spectral features similar to landslides, it can indeed be challenging to accurately distinguish between them. The object-based methods are based on multi-scale or multi-resolution segmentations that use image primitives/features like shape, spectrum, and texture [7]. These techniques consider threshold criteria and hence require appropriate empirical settings for correct results. In addition, the rapid segmentation of high-resolution satellites across large geographical areas is a tough process [8,9]. A comparative study aiming to explore the strengths and weaknesses of pixel and object-based methods for landslide detection has also been suggested in the past [10].

Furthermore, AI methodologies have undergone significant evolution, substantially impacting various domains, including hazard analysis. The emergence of remote sensing-oriented big data has notably enhanced AI's capabilities [11,12]. Machine learning, a sub-field of AI, has shown considerable progress in landslide prevention and assessment [13]. Notably, algorithms such as support vector machines [13–15], logistic regression [16], random forest [17], and decision trees [18] have been applied successfully. While these methods improve efficiency, they often face challenges with accuracy in complex terrains and may require extensive preprocessing. Recent advances in computational technologies have propelled deep learning (a subset of machine learning) into the spotlight, particularly in applications requiring image segmentation [19], object detection [20], and classification [21]. Deep learning has also been increasingly applied in geohazards assessment [4], encompassing earthquakes [22], floods [23], and landslides [24,25].

### **1.1 Related Work**

Diverse architectures of CNNs have been deployed for the detection of landslide-related information [3,4,24,26–28], with notable examples including U-Net [29–33], ResNet [34,35], and Mask-RCNN [36,37]. These models use image patches of multiple sizes for binary segmentation, i.e., landslide and background, yet they face notable challenges. The model by Bragagnolo et al. (2021) struggles to accurately detect landslides in complex terrains where visual features resemble surrounding landscapes, making it difficult to distinguish landslides from other elements [29]. Similarly, Ghorbanzadeh et al. (2021) find it challenging to differentiate between landslides and visually similar bare land in optical imagery, suggesting that SAR data could enhance accuracy by providing additional insights [30]. Devara et al. (2024) highlight the dependence on high-quality training data and the need for specific threshold settings, adding complexity to the detection process [33]. Other studies [2,36,37] have evaluated models using UAV-acquired images, Liu et al. (2024) emphasize the importance of empirical settings for better results when dealing with images of varying sizes [37].

YOLO models [38] specifically, YOLOv3 [39,40], YOLOv4 [41], YOLOv5 [42], YOLOv6 [43], YOLOv7 [44], YOLOv8 [45], and YOLOX [46] have been applied by researchers with significant success. To further refine landslide detection accuracy, attention models have been combined with CNNs [5]. These include 3D-SCAM [47], U-Net + CBAM [48], LA-YOLO-LLL [49], LD-YOLO: based on ECA [50], YOLO-SA [51], SW-MSA [52], YOLOv7-SE [53], LS-YOLO [54], and LA-YOLO (an improved YOLOv8) [55].

### **1.2 Research Gap and Contribution**

Although previous studies have laid a significant foundation, several research gaps remain unsolved. Firstly, YOLOv5 has been extensively applied in other object detection tasks, however, there is limited documentation on its application for landslide detection, particularly in complex terrains using remote sensing data. Existing YOLOv5 studies may lack a thorough evaluation of its performance in detecting landslides under challenging conditions such as varying land cover and cloud cover. Secondly, the efficacy of attention mechanisms within single-stage detection algorithms warrants comprehensive evaluation. Previous efforts integrating attention mechanisms with YOLOv5 often involved altering backbone architectures, making it difficult to isolate and evaluate the direct impact of attention mechanisms on performance. This indicates that the true potential of attention models was not fully explored. Thirdly, identifying an appropriate deep-learning model for accurate landslide detection is imperative. Our study addresses these gaps by adding attention mechanisms within the YOLOv5 architecture without altering other components. This allows us to isolate and fully assess how these mechanisms enhance detection accuracy, particularly in challenging landslide detection tasks where distinguishing landslides from complex backgrounds is crucial. By focusing on relevant image features, attention mechanisms improve the model's precision in

detecting landslides through remote sensing imagery, and our work demonstrates these advances through comprehensive experimentation.

Landslides often occur suddenly and have severe impacts, necessitating timely and precise information for effective emergency actions. Hence, our work contributes to the ongoing development of algorithms that can support real-time emergency response efforts. While other models have shown potential for landslide detection, our research focuses on improving the precision and efficiency of detecting landslides within image patches, utilizing the YOLOv5 + attention model. This specificity in improving detection in challenging scenarios is the unique contribution of our work. By addressing these more detailed aspects of algorithm improvement, our study contributes to the broader goal of rapid and accurate landslide detection, a crucial component of the emergency response framework.

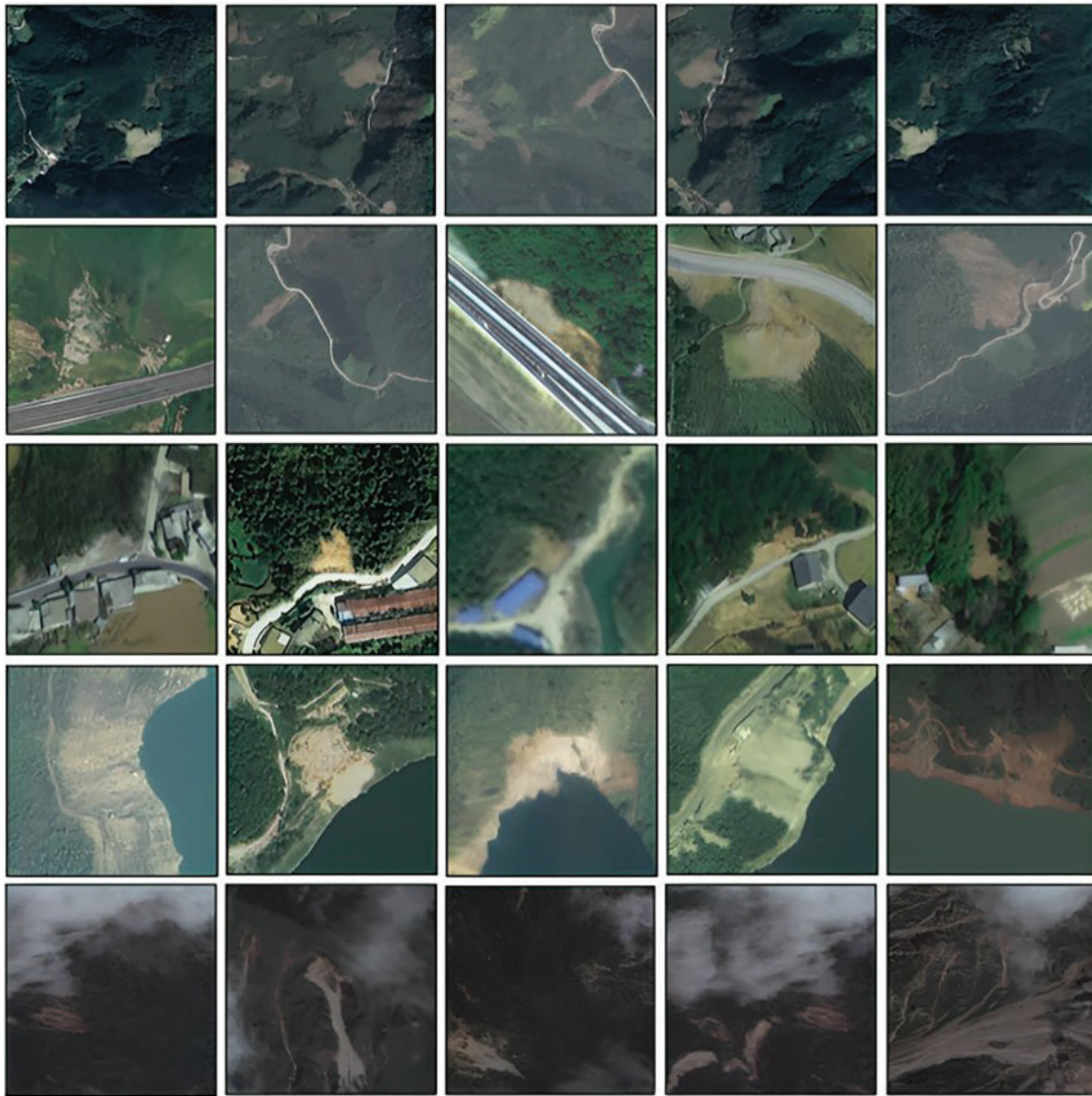
The novel contributions of our work are threefold:

1. The development and optimization of improved YOLOv5 tailored for automated landslide event detection. We enhance YOLOv5 by refining its architecture to improve feature extraction and classification accuracy in landslide-prone regions.
2. The integration of cognitive capabilities through attention mechanisms, specifically CBAM and ECA, into the YOLOv5 architecture to improve hazard scenarios. We embed attention modules within the YOLOv5 backbone to refine spatial and channel-wise feature representation, emphasizing discriminative landslide-related patterns.
3. A comprehensive assessment of the model, benchmarked against satellite datasets encompassing diverse geomorphological characteristics. The results validate the effectiveness of our approach, demonstrating superior performance over conventional methods.

## 2 Study Area and Data Set Description

The research encompasses study sites within Beichuan County in Sichuan Province, Bijie City (Guizhou Province), and Ludian County in Yunnan Province. These areas are recognized for their susceptibility to natural disasters, particularly earthquakes and landslides, which have historically resulted in extensive infrastructural damage [5]. The selected locations are characterized by high mountain ranges, expansive terrain, and dense river networks. The co-seismic landslide events in these regions have inflicted significant damage on essential infrastructure, including tunnels, reservoirs, bridges, roads, and agricultural lands [5]. For this study, a landslide dataset comprising three-channel (RGB) images was acquired from TripleSat (spatial resolution = 0.8 m) for Bijie City [47]), covering an area of 26,853 km<sup>2</sup>, captured from May to August 2018 while imagery for Beichuan and Ludian Counties was obtained from the 91 wemap platform [5]. This database comprises a total of 950 images, each with a resolution of 512 × 512 pixels, accompanied by reference data. Fig. 1 exhibits sample images from the dataset, showcasing the diverse landforms, such as vegetative cover, roadway networks, built-up areas, and aquatic features. The presence of cloud cover within the dataset presents additional complexity, making it particularly challenging for landslide detection algorithms. The dataset can be downloaded from <https://github.com/Abbott-max/dataset> (accessed on 01 January 2025).





**Figure 1:** Examples of the dataset used for experiments [5]

### 3 Methodology

The methodology proposed in this work delineates a systematic procedure for automated landslide detection utilizing satellite imagery. The approach is structured into four primary stages: (i) data annotation and preparation, (ii) model development and architectural design, (iii) training procedures, and (iv) evaluation metrics. The procedural framework of the proposed attention-enhanced YOLOv5 model is encapsulated in Algorithm 1.

**Algorithm 1:** Pseudo-code of attention-based YOLOv5 network

---

Input: Satellite imagery of landslide events, labels\_xml.

Output: Detected bounding boxes and coordinates.

1. Annotation setting: Conversion into YOLO based format.
  2. Data configuration and preparation: Training, validation, and test sets.
  3. Model training: YOLOv5 (n, s, and m), and YOLOv5 + attention (CBAM, and ECA).
  4. *For each epoch do.*
  5. Hyperparameter settings (batch size, image size, data augmentation, etc.).
  6. Estimate the loss function and fine-tune the parameters.
  7. Validation: Evaluating model detection performance using the validation set.
  8. *end for.*
  9. Save bestweight.pt.
  10. Model testing: Trained YOLOv5 + attention to extract landslides from the test images.
  11. Predicted bounding boxes.
- 

**3.1 Input Data, Annotation, and Formation**

The input datasets for our study consist primarily of high-resolution satellite images described in [Section 2](#). The initial annotations of the dataset are represented in the form of polygons (file\_type = xml). Therefore, we transformed the data into the required format, i.e., bounding boxes, comprising five parameters illustrated in Algorithm 2.

**Algorithm 2:** Annotation conversion

---

Input: Folder path with XML annotation files (*xml\_folder*), Output directory (*output\_dir*).

Output: Annotated files in YOLO format.

1. function CONVERT\_TO\_YOLO (*xml\_folder*, *output\_dir*)
  2.   for all *xml\_file* in *xml\_folder* do
  3.     *image\_name*, *width*, *height*, *annotations*  $\leftarrow$  *parse\_xml* (*xml\_file*)
  4.     *annotation\_file*  $\leftarrow$  *open* (*output\_dir* + '/' + *image\_name* + '.txt', 'w')
  5.   For all *annotations* in *annotations* do
  6.     *class\_label*  $\leftarrow$  *annotation*['name']
  7.     *x\_min*  $\leftarrow$  *annotation*['bndbox']['xmin']
  8.     *y\_min*  $\leftarrow$  *annotation*['bndbox']['ymin']
  9.     *x\_max*  $\leftarrow$  *annotation*['bndbox']['xmax']
  10.    *y\_max*  $\leftarrow$  *annotation*['bndbox']['ymax']
  11.    *x\_center*  $\leftarrow$  (*x\_min* + *x\_max*)/2
  12.    *y\_center*  $\leftarrow$  (*y\_min* + *y\_max*)/2
  13.    *box width*  $\leftarrow$  *x\_max* - *x\_min*
  14.    *box height*  $\leftarrow$  *y\_max* - *y\_min*
  15.    *x\_center\_norm*  $\leftarrow$  *x\_center*/*width*
  16.    *y\_center\_norm*  $\leftarrow$  *y\_center*/*height*
  17.    *box\_width\_norm*  $\leftarrow$  *box\_width*/*width*
  18.    *box\_height\_norm*  $\leftarrow$  *box\_height*/*height*
  19.    *write\_to\_file*(*annotation\_file*, *class\_label*, *x\_center\_norm*, *y\_center\_norm*, *box\_width\_norm*,  
      *box\_height\_norm*)
- 

(Continued)

**Algorithm 2 (continued)**


---

```

20.   end for
21.     close (annotation file)
22.   end for
23. end function
24. function PARSE_XML (xml_file)
25. annotations ← []
26. parse xml_file and extract image_name, width, height
27. For all object in xml_file do
28. annotation ← {}
29. annotation['name'] ← extract class label from object
30. annotation['bndbox']['xmin'] ← extract xmin from object
31. annotation['bndbox']['ymin'] ← extract ymin from object
32. annotation['bndbox']['xmax'] ← extract xmax from object
33. annotation['bndbox']['ymax'] ← extract ymax from object
34. Append annotation to annotations
35.   end for
36. Return image name, width, height, and annotations
37. end function
38. function WRITE_TO_FILE (file, class_label, x_center_norm, y_center_norm, box_width_norm,
    box_height_norm)
39. file.write (class_label + '.' + x_center_norm + '.' + y_center_norm + '.' + box_width_norm + '.' +
    box_height_norm + '\n')
40. end function

```

---

Additionally, a YAML file (a standard YOLO format), containing necessary information such as the file paths for training, validation, and testing datasets, the number of classes, and the class names (landslide and background), has been created. To facilitate experimentation, the annotated data has been divided into three subsets: 70% for training, 20% for validation, and 10% for testing. In our study, we used the Roboflow-guided approach for this data distribution. Roboflow helps manage and prepare datasets for machine learning by automatically partitioning them into balanced training, validation, and testing subsets. This ensures diverse and representative data distribution, reducing bias and improving the model's accuracy and generalization for landslide detection.

### 3.2 Model Development and Architecture

This section explains the baseline model (YOLOv5), the implemented attention modules (CBAM and ECA), and the proposed network in detail.

#### 3.2.1 Description of the Original YOLOv5 Network

YOLO models [56,57] are popular one-staged object detection algorithms comprising three essential units. First, the Backbone, a CNN, constructs image features. Second, the Neck, a series of layers, combines different image features for subsequent predictions. Third, the **Head** accumulates image features from the preceding module (neck), facilitating predictions. YOLOv5, presented by Ultralytics (<https://github.com/ultralytics/yolov5> (accessed on 01 January 2025)), is a prominent addition to the YOLO family.

We select YOLOv5 as our primary model for three main reasons. First, YOLOv5 integrates Cross-Stage Partial Network (CSPNet) within the Darknet, forming CSPDarknet, the backbone [58]. CSPNet addresses issues related to redundant gradient information by incorporating gradient changes into the feature map [58]. This approach reduces the model's parameters and Floating-Point Operations Per Second, ensuring both inference speed and accuracy while minimizing model size as well as the significance compact model size for efficient inference on restricted computational platforms. YOLOv5 with CSPDarknet is a suitable choice.

The main architecture consists of stacking multiple CBS (convolution + batch normalization + sigmoid linear unit) nodes and concentrated-comprehensive-convolution (C3), modules, with one spatial-pyramid-pooling-fast (SPPF) module connected at the end [59]. Second, YOLOv5 uses the Path Aggregation Network (PANet) [60] to enhance information flow. PANet employs a Feature Pyramid Network (FPN) [61] with an enhanced bottom-up path, facilitating the propagation of low-level features and aiding the model in generalizing effectively to objects of various sizes and scales. These significant characteristics of the neck module of YOLOv5 make it suitable for complex scenarios like landslide detection. Additionally, adaptive feature pooling is utilized [60], connecting the feature grid with all feature levels to enable useful information in each level to propagate directly to subsequent sub-networks. PANet enhances the utilization of precise localization signals in lower layers, thereby significantly refining object location accuracy [60].

Thirdly, the YOLOv5 head generates feature maps of three different sizes to facilitate multi-scale prediction [56]. This feature enables the model to detect objects of varying sizes. For example, in landslide detection tasks, where satellite images may contain landslide events of different sizes (including small, medium, and large), multi-scale detection ensures that the model can accurately identify and extract these events regardless of their scale. Further, data augmentation is a key training process in the YOLOv5 model that introduces transformations to the initial training set, expanding the model's exposure to a broader spectrum of semantic variations beyond the isolated training dataset. In particular, mosaic, translation, scaling, flipping, and HSV augmentation techniques are applied in this study to diversify the training data, ensuring the model's capacity to generalize across a range of inputs [62].

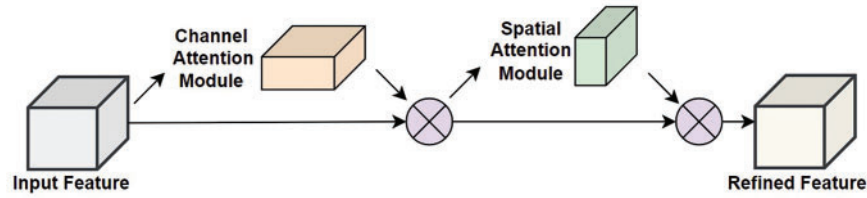
YOLOv5 computes CIoU (complete-intersection-over-union) bounding box regression loss function to optimize the accuracy [63]. CIoU loss incorporates the geometric relationships between predicted and ground truth bounding boxes, leading to better localization accuracy compared to traditional Intersection over Union (IoU). Moreover, CIoU loss considers the aspect ratio and size differences between predicted and ground truth boxes, making it more robust to variations in object size and shape. In addition, CIoU loss effectively accounts for overlapping regions, enabling the model to more accurately distinguish between closely located objects. Compared to the other loss functions, CIoU is less affected by vanishing gradients and saturation, leading to more stable training dynamics and faster convergence [59]. Besides, CIoU loss correlates well with evaluation indicators, ensuring that the loss function aligns with the desired model performance metrics. Furthermore, five versions of YOLOv5 (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x) are available [57]. These versions are initially trained on the large-scale MS-COCO database, enabling them to acquire rich and generalizable features from a diverse array of images. Therefore, when these pre-trained models are used for applications like landslide detection, where collecting large amounts of labeled data may be challenging, their generalization ability allows them to perform well even when trained on relatively small or limited datasets. Considering the size of the dataset and the existing computing facilities, our study focuses on utilizing YOLOv5n, YOLOv5s, and YOLOv5m models.

### 3.2.2 CBAM

According to Woo et al. (2018), a CBAM is an effective attention framework that can be fused into deep learning models with remarkably fewer parameters [64]. As seen in Fig. 2, the CBAM includes two blocks:



(i) a channel attention module (CAM) that undertakes two essential tasks: global average pooling (GAP) followed by maximum global pooling. (ii) A spatial attention module (SAM) that performs maximum as well as average pooling. This architecture enables the computation of attention-based weights for accurate feature map refinement. GAP computes the average value of individual feature maps through all spatial locations, capturing the overall distribution of information, while maximum pooling identifies the most significant features within each feature map by selecting the maximum value. Therefore, the network can capture both the general distribution of features and the most salient features in each map. Further, GAP provides a robust representation of the entire feature map, ensuring that important information is not lost during pooling. Meanwhile, maximum pooling focuses on extracting the most relevant features, enhancing the discriminative power of the pooled representation. Combining these pooling methods creates a comprehensive representation that incorporates global context and local features. The integration of GAP and maximum pooling also allows the network to generalize well to different input variations. GAP provides a stable representation that is less sensitive to small changes in the input, however, maximum pooling focuses on extracting robust features that are invariant to translation or rotation. This combination enriches the network's capability to learn discriminative features across diverse input samples. Hence, combining GAP and maximum pooling enables the network to effectively capture both global context and local characteristics, leading to improved performance in critical tasks like landslide information extraction.



**Figure 2:** Structural framework of CBAM [60]

Furthermore, the mathematical description of CBAM is as follows: via convolution and pooling procedures, the CBAM calculates the 1-D channel attention ( $M_{CA} \in R^{C \times 1 \times 1}$ ) and 2-D spatial attention maps ( $M_{SA} \in R^{1 \times H \times W}$ ) using an input feature map ( $FM \in R^{C \times H \times W}$ ). The total computed attention is estimated by Eqs. (1) and (2).

$$FM' = M_{CA} (FM) \otimes FM \quad (1)$$

$$FM'' = M_{SA} (FM') \otimes FM' \quad (2)$$

where ' $\otimes$ ' denotes element-wise multiplication and  $Fm''$  signifies the enhanced outcome.

In addition, the weights (CAM and SAM) (Woo et al., 2018; Yang et al., 2022) are computed by Eqs. (3) and (4).

$$CAM_{Weights} = Sigmoid(Multilayerperceptron(Maxpooling(FM)) + Multilayerperceptron(Avgeragepooling(FM))) \quad (3)$$

$$SAM_{Weights} = Sigmoid \quad (4)$$

### 3.2.3 ECA-Net

ECA-Net [65], a lightweight model comparable to SENet [66]. ECA-Net is designed to be more computationally efficient than SENet. Due to its squeeze-and-excitation operations, SENet introduces additional parameters and computational overhead, which involve learning parameters for each channel and performing GAP operations. The ECA module introduces minimal computational overhead by performing 1D convolution along the channel dimension, resulting in a more efficient architecture. ECA-Net follows sparsity in the connectivity pattern of the attention mechanism to reduce computational complexity further. This sparsity helps limit the number of connections and operations required for feature recalibration, leading to faster inference and reduced memory consumption compared to SENet. The ECA module effectively captures channel-wise dependencies and adaptively recalibrates features, leading to improved feature representations and enhanced model performance. Thus, ECA-Net is well-suited for environments with limited resources and real-time applications, for example, landslide detection, where efficiency is crucial. ECA-Net offers greater flexibility in terms of model architecture and scalability. The output of the convolutional layers is a 4-D tensor, serving as input to ECA-Net. The architecture of ECA-Net comprises three modules [61]: (i) global feature descriptor, (ii) adaptive neighborhood interaction, and (iii) broadcasted scaling. The GAP operation implicates processing the served input tensor by averaging all pixels in a specific feature map, resulting in a single pixel. Subsequently, the tensor ( $C \times 1 \times 1$ ), undergoes a 1-D striding convolution with a kernel size denoted as ' $K_S$ '. Additionally, the adaptive estimation of ' $K_S$ ' is determined based on the channel space ( $C$ ), utilizing Eqs. (5) and (6).

$$K_S = \psi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{z}{\gamma} \right\rfloor_{odd} \quad (5)$$

$$C = \varnothing(k) = 2^{(\gamma * k - z)} \quad (6)$$

where  $\gamma$ , ' $z$ ' signify the predefined hyper-parameters.

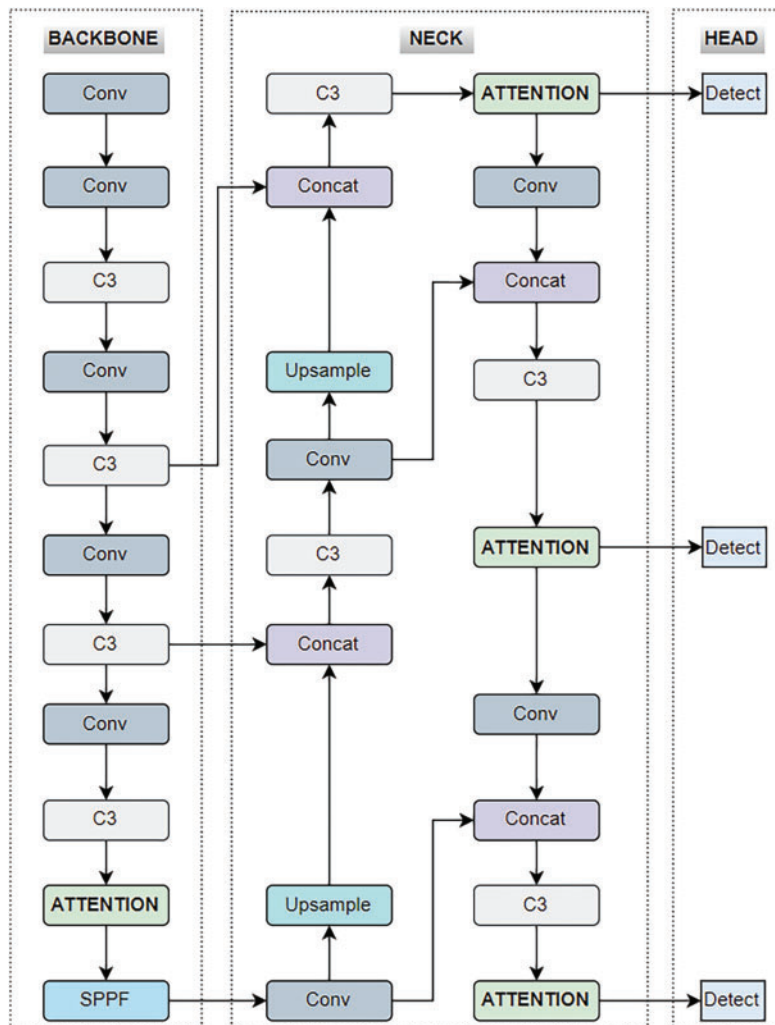
### 3.2.4 Proposed Attention Embedded YOLOv5 Network

Attention, a primary cognitive capability of humanoids, allows them to process the most pertinent information, eliminating the insignificant information. In this regard, attention modules, inspired by the human vision system, show substantial potential by their integration within a CNN to augment its performance. Hence, we amalgamated two attention mechanisms particularly CBAM and ECA, separately in the backbone and the neck of the YOLOv5 network.

The original YOLOv5 backbone encompasses a sequence of convolutional layers, bottleneck blocks (C3, denoting concentrated-comprehensive-convolution), and SPPF (spatial-pyramid-pooling-fast), systematically downsampling the input image. The backbone is optimized by adding an attention module (especially CBAM and ECA discretely) between the C3 layer and SPPF. This augments the network's ability to capture multi-scale spatial and contextual features before they are fed into the neck for further processing. Subsequently, within the neck, the attention module is integrated after the last three C3 layers for enhancement. This ensures feature refinement before multi-scale feature fusion, improving the ability of the PANet structure to focus on important landslide features.

The architecture of the proposed network is shown in Fig. 3. Embedding attention in YOLOv5n, YOLOv5s, and YOLOv5m networks aids several parameters. First, model depth\_multiple, layer channel\_multiple, and the total number of classes (01). The model's depth is regulated by the depth\_multiple parameter, which scales the number of layers within each module (C3 or Conv). Acting as a multiplier, it plays

a key role in determining the overall depth of the network. However, the `channel_multiple` parameter serves to scale the number of channels in each layer, thereby influencing the model's width. This critical parameter directly impacts the computational settings and the memory requirements, shaping the trade-off between model complexity and resource efficacy. Second, it creates anchor boxes based on various detection scales within the convolutional layer of its backbone. These anchor boxes are vital for ascertaining the dimensions and positions/locations of the detected objects. The addition of an attention module in the final layer of the backbone enhances the YOLOv5 discriminative capabilities. Moreover, the neck defines the construction of the detection head in YOLOv5, comprising convolutional layers, upsampling layers, and concatenation tasks. These elements assist in the integration of feature maps derived from different scales of convolutional layers within the backbone of the proposed network. Further, an attention module is appended at the end of each layer to refine the extracted features. Based on the generated feature maps and anchor boxes, the head (detection module) performs the task of extracting landslide events. The proposed architecture is aimed at enhancing accuracy, enabling real-time extraction of landslide events from remote sensing data.



**Figure 3:** Proposed YOLOv5 + attention-based network for landslide detection

### 3.3 Training Setup

During training, an image size of  $640 \times 640$  pixels is served as an input to the proposed network. The choice of a  $640 \times 640$  pixel image size for our landslide detection model is strategically aligned with the spatial resolution of the imagery and the scale of landslide features. This image size is generally sufficient to capture a variety of landslide sizes, ensuring that the model can detect features effectively while maintaining a balance between detail and computational efficiency. Additionally, the YOLOv5 architecture is designed to efficiently process input images of size  $640 \times 640$  pixels. Moreover, the batch size is 8, epochs 500. We used a stochastic gradient descent optimizer with an initial learning rate of 0.01, an initial momentum factor of 0.937, and an initial weight decay of 0.0005. Besides, the considered hyperparameters are based on previous studies [44,46] and directly adopted for the current experiment. Furthermore, the machine configuration includes CPU: Intel® Xeon-CPU E3-1231 (v3@3.40 GHz), GPU (graphics processing unit): NVIDIA-GeForce-RTX3080Ti, deep learning framework: PyTorch 1.7, operating system: Ubuntu 18.04, and CUDA11.4.

### 3.4 Evaluation Criteria

The performance of the YOLOv5 + attention model is evaluated using both quantitative and qualitative criteria. The four primary metrics used for quantitative assessment are precision, recall, F-score, and mAP (Eqs (7)–(10)). The use of these is well-established and has been extensively used in previous studies [5,48,49,67,68]. Below is a detailed explanation of each metric:

$$Precision = \frac{(Truepositive)}{(Truepositive) + (Falsepositive)} \quad (7)$$

$$Recall = \frac{(Truepositive)}{(Truepositive) + (Falsenegative)} \quad (8)$$

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

$$mAP = \sum_{i=0}^n \frac{1}{n} \int_0^1 Precision(Recall) d(Recall) \quad (10)$$

On the other hand, qualitative evaluation is conducted through visual analysis of the computed results.

## 4 Experimental Results

Here, the experimental findings of the proposed network are given. First, the quantitative results derived from the metrics explained in Section 3.4 are presented. Second, the qualitative outputs of the extracted landslides are described. Third, the comparative analysis of the obtained results is described. Fourth, the overall discussion is mentioned.

### 4.1 Quantitative Evaluation

The results of quantitative indicators, i.e., mAP, precision, recall, and f-score of the attention-embedded YOLOv5 are shown in Table 1. The mAP@0.5 of the YOLOv5 variants (baseline) particularly, YOLOv5n, YOLOv5s, and YOLOv5m is 73.1%, 75.1%, and 72.5%, respectively.

**Table 1:** Experimental results of YOLOv5 + attention-based models

Model	GFLOPs	Parameters	Precision	Recall	F-Score	mAP@0.5	Time	GPU
	(m)	(m)					(h)	(GBs)
YOLOv5n	4.1	1.76	71.3	74.5	72.9	73.1	0.54	1.10
YOLOv5n + CBAM	4.3	1.91	84.4	71.1	77.2	<b>77.8</b>	0.88	1.10
YOLOv5n + ECA	8.1	3.91	76.0	75.2	75.6	77.0	1.19	1.93
YOLOv5s	15.8	7.01	73.3	67.4	70.2	75.1	0.53	1.88
YOLOv5s + CBAM	16.2	7.62	70.4	73.9	72.1	76.1	0.68	2.17
YOLOv5s + ECA	31.5	15.61	80.7	70.5	75.3	76.8	1.53	3.88
YOLOv5m	47.9	20.85	75.2	64.7	69.6	72.5	0.91	3.73
YOLOv5m + CBAM	49.0	22.21	76.4	75.3	75.8	76.8	1.33	3.85

The mAP is a comprehensive metric that summarizes the precision-recall trade-off across different confidence thresholds. An increase in mAP indicates an overall improvement in the model's ability, providing a more accurate and reliable detection performance across various conditions. A higher mAP reflects that the model is consistently detecting landslides with both high precision and high recall, making it effective in diverse and challenging environments. Considering the YOLOv5n + CBAM model, the mAP@0.5 computed is 77.8%, which showed an improvement of +4.7%. Further, an enhancement of +1.0% is noticed in mAP@0.5 of YOLOv5s + CBAM (76.1%). Moreover, in the case of YOLOv5m + CBAM, the estimated progress is +4.3% (mAP@0.5). Besides, ECA assimilated YOLOv5 also demonstrated the increased accuracy, mainly YOLOv5n + ECA (mAP@0.5 = 77.0%) and YOLOv5s + ECA (mAP@0.5 = 76.8%) unveil the progress of +3.9%, and +1.7%, respectively. Here, mAP evaluates how well the YOLOv5 model, enhanced with attention mechanisms (CBAM and ECA), detects landslides across a range of confidence thresholds. Unlike precision and recall, which are evaluated at a single threshold, mAP considers the performance at multiple thresholds, providing a holistic view of the model's accuracy. In landslide detection, variations in confidence levels can affect the reliability of detections. mAP provides a comprehensive summary of how well the model performs across these different levels, ensuring that the evaluation is not biased by a single confidence threshold.

In addition, a difference of 1.60% is observed in the f-score of YOLOv5n + CBAM and YOLOv5n + ECA; however, the improvement reached +4.3% and +2.7%, respectively. Significant progress of +1.9% and +5.1% is seen in the f-score of YOLOv5s + CBAM and YOLOv5s + ECA, respectively. The highest improvement of +6.2% in the F-score is obtained in the case of YOLOv5s + CBAM. The F-score is a critical metric for evaluating landslide detection models as it balances precision and recall, handles imbalanced datasets, provides a single performance metric for easy comparison, helps interpret model trade-offs, and reflects real-world applicability. For landslide detection using models like YOLOv5 enhanced with attention mechanisms, the F-score ensures that both false positives and false negatives are minimized, leading to more accurate and reliable detection and mapping of landslides. The precision and recall ranged from 70.4% to 84.4% and 64.4% to 75.3%. Precision measures the proportion of correctly identified landslides out of all detected landslides. However, recall estimates the proportion of actual landslides that the model successfully detects. Considering the evaluation metrics, the YOLOv5n + CBAM exhibited the best performance.

In landslide detection tasks, calculating GLOPs (Giga Floating Point Operations), the number of parameters, the GPU used, and computational time are essential for estimating the model's performance.



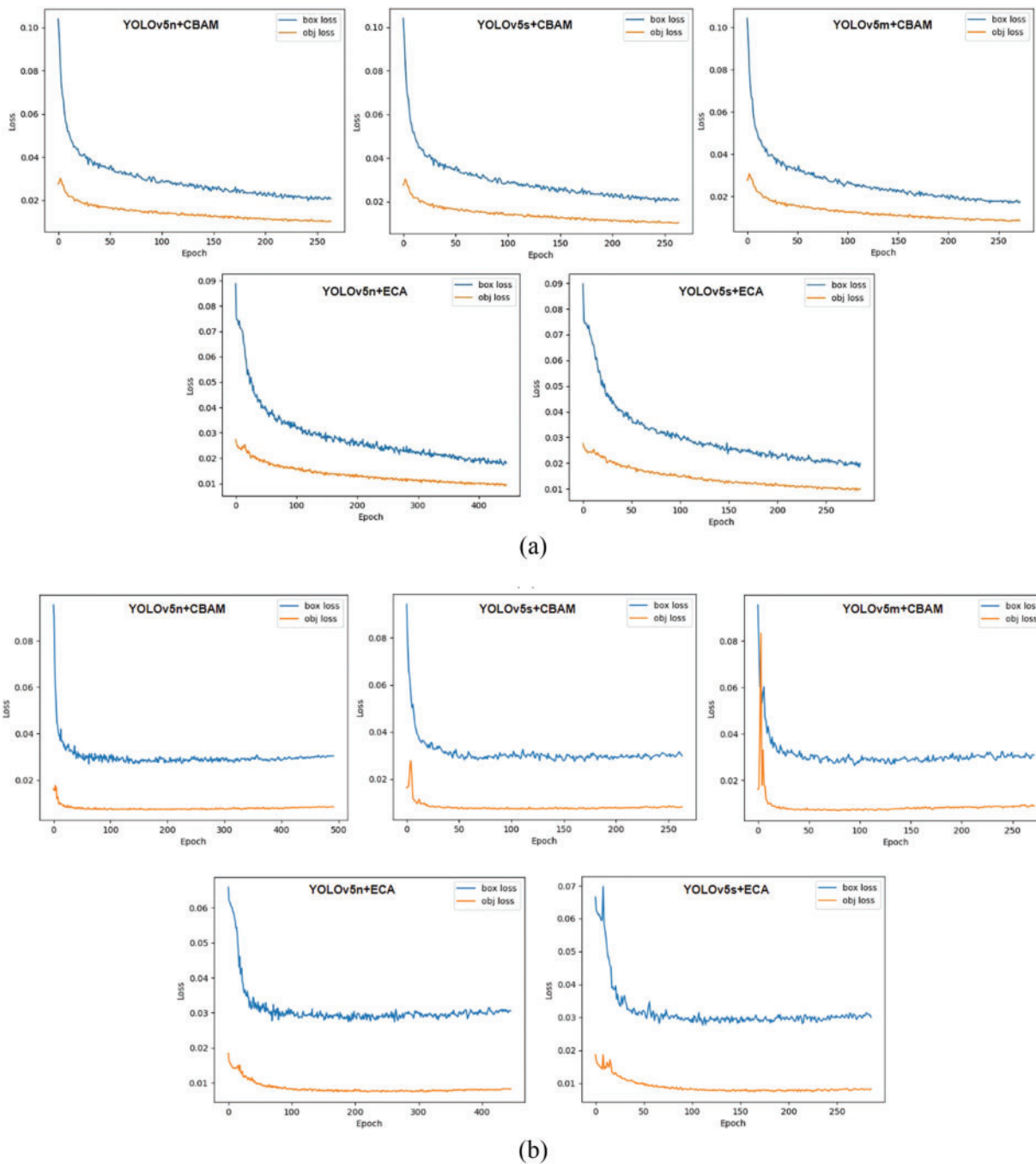
GLOPs and parameters provide insights into the model's complexity and computational load, which directly affect its efficiency and feasibility for real-time applications. The computed GLOPs (highest = 49.0 (YOLOv5m + CBAM), and lowest = 4.1 (YOLOv5n)) and the number of parameters (maximum = 22.21 (YOLOv5m + CBAM), and minimum = 1.76 (YOLOv5n)) in millions (m) by the baseline and attention integrated models are also shown in [Table 1](#).

The computational time is an important parameter for practical deployment, especially in disaster response scenarios where timely detection and mapping of landslides are imperative. Together, these metrics help in evaluating the model's performance, optimizing resource allocation, and ensuring that the system can deliver rapid and accurate results, thereby enhancing its applicability and reliability in real-world landslide detection tasks. The execution time of the YOLOv5s + ECA network is maximum (1.53 h) while the YOLOv5n + CBAM takes the minimum time (0.88 h). The execution time of two baseline models (YOLOv5n = 0.54 h and YOLOv5s = 0.53 h) is almost similar (difference = 0.01 h), whereas YOLOv5m is executed in 0.91 h.

Our study is limited to analyzing landslides within fixed-size image patches, and as such, the total coverage area is not explicitly defined in our current analysis. While this patch-based approach simplifies detailed analysis, it doesn't assess performance over larger, contiguous regions, which would require processing multiple patches for broader applications.

Furthermore, the GPU used is crucial as it determines the model's processing capability and speed; different GPUs offer varying levels of performance that can significantly impact the model's ability to handle large datasets quickly. The GPU or memory utilization (GBs) by each model is also given in [Table 1](#). It is observed that GPU consumption is directly related to the size of the model (YOLOv5n = 1.10 GBs, YOLOv5s = 1.88 GBs, and YOLOv5m = 3.73 GBs). Moreover, estimating the loss functions is crucial for the proposed landslide detection. The bounding box regression loss computes how accurately the predicted bounding boxes align with the actual landslide locations, ensuring that the model precisely identifies the spatial extent of landslides. Confidence loss, on the other hand, evaluates the model's certainty in its predictions, distinguishing between true landslide detections and false positives.

The assessed loss functions (training and validation) of the five proposed networks (YOLOv5n + CBAM, YOLOv5s + CBAM, YOLOv5m + CBAM, YOLOv5n + ECA, and YOLOv5s + ECA) are presented in [Fig. 4a,b](#), respectively. When the bounding box regression loss is low, it indicates that the model is accurately predicting the locations of the landslides with minimal error, i.e., the predicted boxes are close to the actual ground truth boxes. However, a high bounding box regression loss suggests that the predicted bounding boxes are not aligned well with the true locations of the landslides. This could indicate that the model struggles to localize the landslide accurately, resulting in poor detection performance. Further, confidence loss estimates how assured the model is in its prediction that a detected object is indeed a landslide. A low confidence loss means the model is effectively distinguishing between landslides and non-landslide regions. The model is confident when it makes correct detections and assigns high confidence scores to these predictions. Higher confidence loss shows that the model is either missing some landslides (false negatives) or incorrectly identifying non-landslide regions as landslides (false positives). This reflects a lack of confidence in making accurate predictions. Hence, the lower function represents that the model is performing well (YOLOv5n + CBAM). It can accurately locate and identify landslides with high confidence, meaning that both the bounding box and object classification tasks are being handled correctly. On the other hand, a higher loss function signifies that the model needs further training or optimization. Either it struggles with localizing landslides, identifying them, or both, leading to errors.

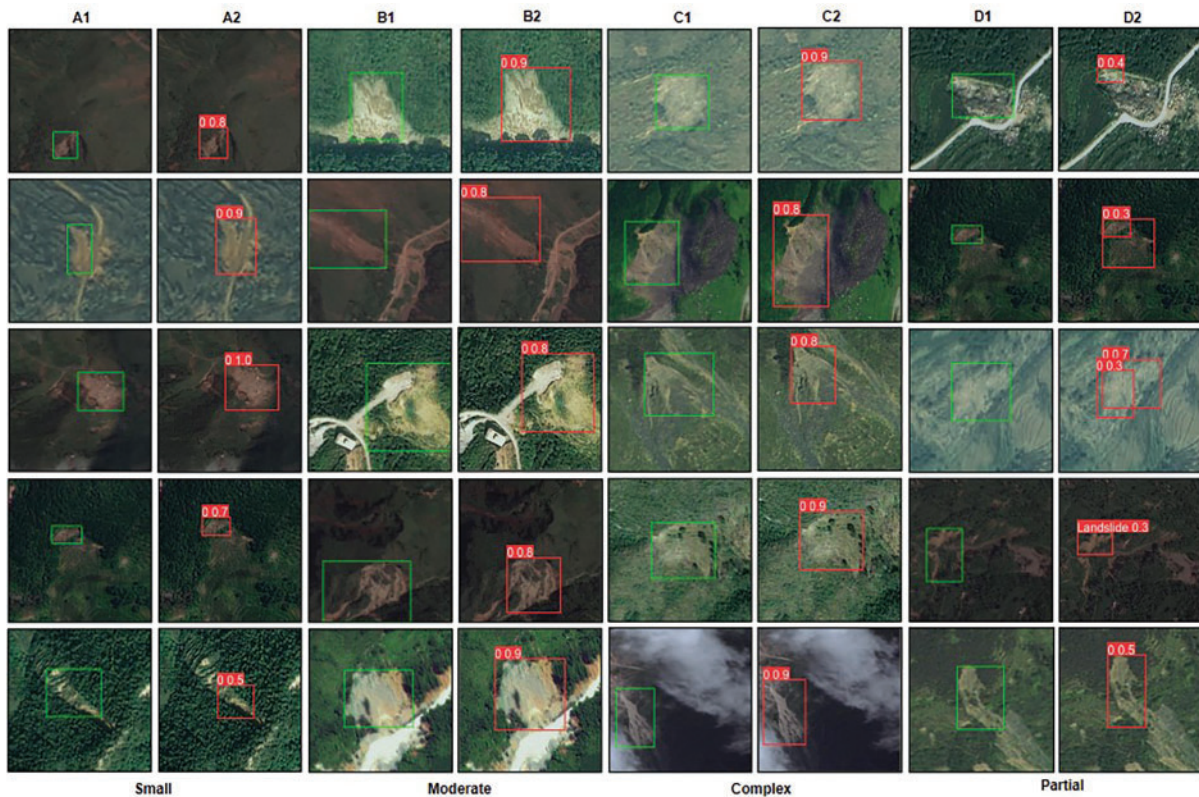


**Figure 4:** (a). Training loss function estimation. (b). Validation loss function estimation

#### 4.2 Qualitative Evaluation

The visual outcomes of the extracted landslides by YOLO plus attention models are represented by bounding boxes shown in Fig. 5. The bounding box is a standard in object detection frameworks, especially those based on the YOLO architecture. Bounding boxes provide a simplified and efficient means of detecting and localizing objects within an image. The primary goal of the YOLOv5 + attention model is to quickly identify potential landslide areas from satellite imagery. Bounding boxes allow for rapid detection, which

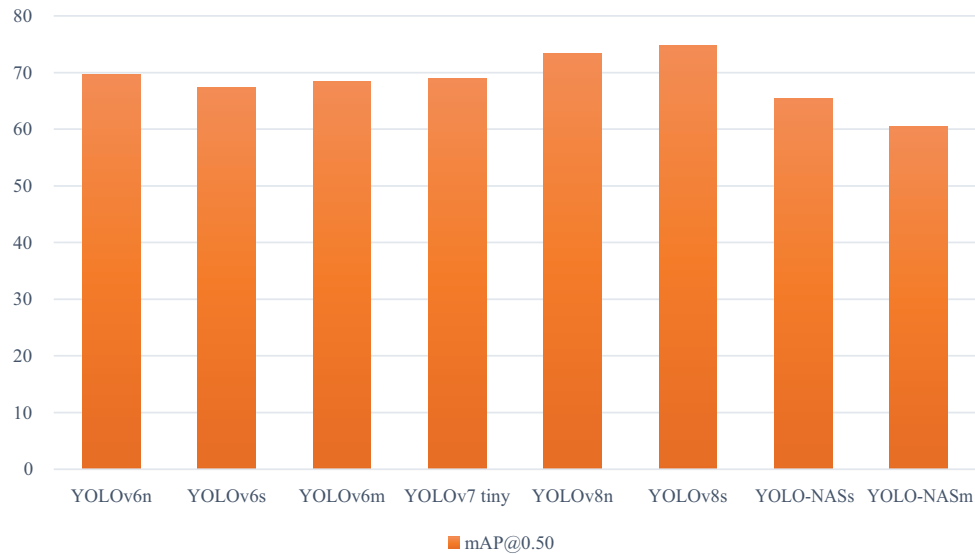
is crucial for time-sensitive applications such as post-disaster response. Moreover, training a model to detect and delineate exact polygonal outlines is significantly more complex and computationally intensive compared to detecting bounding boxes. Bounding boxes serve as a practical compromise that enables efficient training and inference while still providing valuable spatial information about the landslides. The use of bounding boxes allows the model to scale more effectively when processing large geographical areas. Given the extensive size of satellite images and the potential number of landslide events to be detected, bounding boxes enable the model to process images more quickly and with fewer computational resources. While bounding boxes provide a useful starting point for landslide detection, we acknowledge that detailed polygonal outlines are more desirable for many applications, such as accurate mapping and geospatial analysis. Irrespective of the spatial location, the landslides of small (A1), as well as moderate (B1) sizes in an image, are successfully extracted by the YOLOv5n + CBAM model. The model can differentiate the landslides of different extents. The dataset comprises images with complex backgrounds, which makes it challenging for landslide detection. For example, the presence of clouds, and landscapes with homogenous characteristics (landslides, and bare sand). In Fig. 5 C1, some of the landslides are difficult to interpret manually, however model predicts them effectively. The images in column 4 (D1) indicate the examples where the models struggle to detect landslide events. Nevertheless, YOLOv5n + CBAM showcased high detection precision and robustness. Hence, automated landslide recognition models are anticipated for better prediction results.



**Figure 5:** Qualitative results of the extracted landslides by YOLOv5n + CBAM (Labels: A1, B1, C1, and D1, detected landslides: A2, B2, C2, and D2)

### 4.3 Comparison with the Advanced YOLO Models and Previous Work

For the considered dataset, the competency of the advanced YOLO models specifically, YOLOv6 (n, s, and m), YOLOv7 tiny, and YOLOv8 (n, and s) is evaluated. Moreover, YOLO-NAS (s, and m), a recent object detection model, is also executed to examine its capability in the studies of landslide hazards. In total, we compared the obtained results with eight state-of-the-art YOLO models. The obtained mAP@0.5 of all the models (Fig. 6). The results indicate that the maximum accuracy is achieved by YOLOv8s (74.8%) followed by YOLOv8n (73.4%). The performance of YOLO-NASm (60.4%) is the least; however, a notable improvement is seen in YOLO-NASs (65.4%). A difference of 1.2% is observed in the case of YOLOv6s (67.3%), and YOLOv6m (68.5%), whereas almost similar accuracy is estimated for YOLOv6n (69.7%), and YOLOv7 tiny (69.0%) models. The significant improvement of +3.0% to +17.4%, and +2.2% to +16.6% in the mAP@0.50 of CBAM, and ECA-integrated YOLOv5 networks, respectively demonstrates the superior performance of our approach compared to other state-of-the-art YOLO models. Our model not only outperforms earlier YOLO versions but even more advanced models such as YOLOv8 and YOLO-NAS. This comparison and wide range of improvement prove the robustness of our attention-enhanced architecture across challenging environments.



**Figure 6:** Performance of the advanced YOLO models using satellite images

Further, the lack of direct comparison with previous studies in our work is primarily due to the unique nature of our dataset and the novel combination of models we employed. The dataset used in our study has not been extensively utilized in previous studies. This uniqueness makes direct comparison challenging. Earlier studies on landslide detection have used different datasets, making direct performance comparison unfeasible. A justified comparison can only be made with studies using the same data and models. Our study aims to establish a baseline for future research using this specific dataset and methodology. However, we compared our results with the existing YOLOv5 + attention models applied to other datasets with dissimilar characteristics. The comparison between the suggested YOLOv5n + CBAM, YOLOv5n + ECA models and earlier studies, particularly LP-YOLO [44] and YOLOv5 + ASSF + CBAM [69], highlights the important advancements made in terms of performance metrics, specifically mAP@0.5, precision, and recall. Compared to LP-YOLO, our YOLOv5n + CBAM model shows a significant improvement in mAP@0.5 (+28.8%), precision (+30.7%), and recall (+21.3%). This shows that our model is enhanced at correctly identifying



landslides, minimizing both false positives and false negatives, which is critical in applications requiring high accuracy. Moreover, YOLOv5n + ECA also shows a noteworthy improvement over LP-YOLO, with increases in mAP@0.5 (+28.0%), precision (+22.3%), and recall (+25.4%). These metrics demonstrate that the ECA-based model also yields challenging results.

In comparison with the YOLOv5 + ASSF + CBAM model [64], our YOLOv5n + CBAM shows a +3.8% increase in mAP@0.5 and an improvement of +6.0% in precision, although no progress is observed in recall. This illustrates that our model not only matches existing attention-based approaches but even exceeds them in terms of accuracy and precision, making it a stronger alternative. In addition, YOLOv5n + ECA also demonstrates an improvement of +3.0% in mAP@0.5, further showcasing how incorporating different attention mechanisms can result in improvements over previously existing models.

#### 4.4 Discussion

The proposed model is a trained YOLOv5 model enhanced with attention mechanisms such as CBAM and ECA. This advanced model is specifically designed for the detection and mapping of landslides using remote sensing imagery. The model's training process involves several critical phases, which include the acquisition and annotation of training data, the integration of attention mechanisms during model training, and the fine-tuning and validation of the model to ensure optimal performance. These steps are essential for developing a model that can accurately identify landslide features in input images and generate detailed, reliable outputs. The timescales for data generation are largely dependent on the size and complexity of the dataset; however, once trained, the model is capable of processing and generating maps for small to medium-sized areas efficiently. The minimum mappable feature size is a critical factor in the effectiveness of our YOLOv5 + attention model for landslide detection. The model's ability to identify and map small landslide features is constrained by the resolution of the input images and the model's detection capabilities. Regarding input requirements, the model needs high-resolution satellite images to effectively detect landslides, with multi-spectral images being preferred due to their ability to capture detailed features. Additionally, high-resolution UAV images can be utilized for detailed analyses of smaller areas. The output from the model is presented as bounding boxes that accurately indicate the presence of landslides, aiming for high spatial accuracy with minimal misalignment. This ensures that the detected landslides closely match their actual locations.

Moreover, a relatively lower recall of the YOLOv5n + CBAM model is noted, despite its high precision and F-score. This indicates that while the model effectively minimizes false positives, it may overlook some actual landslide occurrences. To address this, we analyzed potential reasons for the lower recall. Filtering of subtle features: The CBAM module enhances feature selection but might suppress small-scale landslides. Detection threshold: A high confidence threshold may lead to fewer false positives but could also exclude some true positives. Dataset characteristics: The presence of complex terrain and vegetation may introduce challenges in detecting all landslides. Spatial resolution constraints: Smaller landslides may not be well represented due to the image resolution limitations. To improve recall while maintaining high precision, we plan to explore several approaches. Fine-tuning the detection confidence threshold could help retain more potential detections without significantly increasing false positives. Data augmentation techniques, such as generating synthetic samples and applying transformations, can enhance the representation of underrepresented landslide features. Additionally, incorporating hybrid attention mechanisms or multi-scale feature extraction strategies may allow the model to capture more complex landslide patterns effectively. Besides, post-processing techniques such as morphological operations or region-growing algorithms could be utilized to refine detections and improve recall.



Additionally, from [Table 1](#), we observe that models integrated with attention mechanisms generally enhance detection performance in terms of precision, recall, and F-score, but at the cost of increased GFLOPs, parameter size, and computational time. For instance, YOLOv5n + CBAM improves the F-score and mAP@0.5 compared to the baseline YOLOv5n while maintaining a relatively small increase in computational complexity. This suggests that CBAM enhances feature extraction efficiency with only a small computational overhead. However, a different trend is seen in YOLOv5s + ECA, which achieves a higher precision (80.7%) but at a significant computational cost and a GPU memory demand. This highlights that while ECA-based models improve accuracy, they also require substantially more computational resources. Based on the results, we noted that the YOLOv5n + CBAM model presents a strong balance between performance and efficiency, making it an ideal choice for applications where both accuracy and computational constraints need to be considered. On the other hand, if higher accuracy is the priority and computational resources are not a constraint, models like YOLOv5s + ECA can be beneficial. Furthermore, we discuss the performance of an additional attention mechanism integrated within YOLOv5 to infer its capability. Next, we evaluate the potential of the state-of-the-art YOLO-based models for extracting landslides. Afterward, the implication of the patience parameter in the YOLO model is highlighted.

#### 4.4.1 Additional Attention Mechanism-Based Inferences

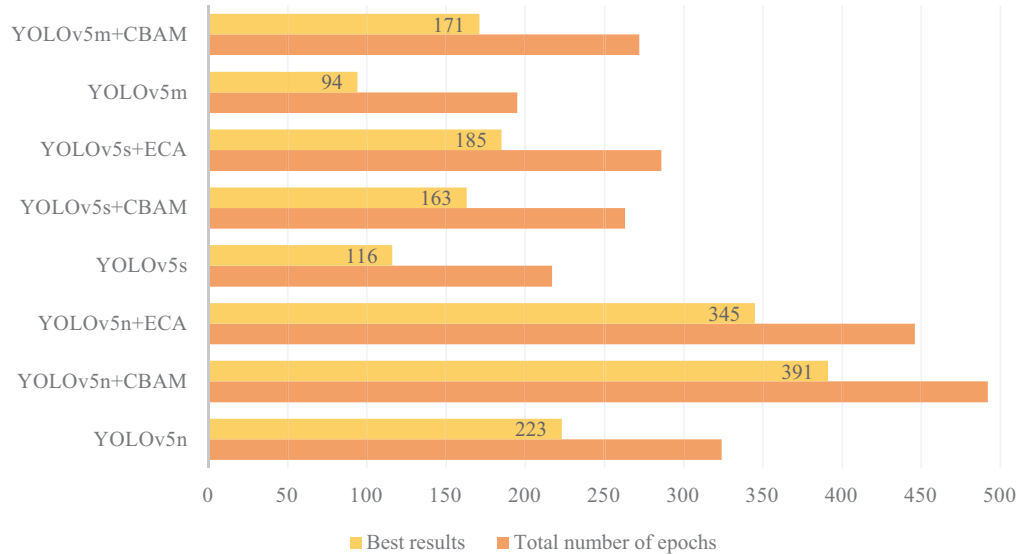
We deliberated on the supplementary attention model to determine the appropriate model for landslide detection. Specifically, GAM (global attention mechanism) [70] restructures the channel and spatial modules of CBAM innovatively. A major improvement (+5.4%) is noted in the f-score (another evaluation indicator) of YOLOv5s + GAM (75.6%), in comparison to the YOLOv5n (70.2%). The computed mAP@0.50 of YOLOv5n + GAM = 74.4%, and YOLOv5s + GAM = 75.9%. Compared with the baseline models (YOLOv5n = 73.1%, and YOLOv5s = 75.1%), progress of +1.3%, and +0.8% is found in the case of YOLOv5n + GAM, and YOLOv5s + GAM, respectively. Additionally, the precision (YOLOv5n + GAM = 73.8%, and YOLOv5s + GAM = 74.9%) and recall (YOLOv5s + GAM = 76.3%) of GAM is improved than YOLOv5n (precision = 71.3%, and recall = 74.5%), and YOLOv5s (precision = 73.3%, and recall = 67.4%). However, the computed recall (71.6%) of YOLOv5n + GAM is less than the baseline model. Furthermore, there was a notable difference of 15.8 GFLOPs between YOLOv5n + GAM (5.5) and YOLOv5s + GAM (21.3), indicating increased computational complexity for the latter. The number of parameters in YOLOv5s + GAM (21.3 million) was also significantly greater than that of YOLOv5n + GAM (5.5 million). Subsequently, the execution time of YOLOv5s + GAM (1.46 h) was higher than that of YOLOv5n + GAM (1.06 h) network, reflecting the increased resource demands of the more complex architecture.

#### 4.4.2 Importance of Early Stopping Conditions

During the experimentation with machine learning and deep learning, early stopping is a method applied to aid the network's efficacy. It involves the constant monitoring of evaluation indicators while training and terminating the training procedure if no enhancement is observed after certain epochs, technically called the patience parameter (which is user-defined). It helps achieve a balance between training time and overall model performance.

Selecting the appropriate patience parameter is key to improving performance. If the patience is set too low, the model training might end prematurely, potentially overlooking possible performance enhancements. Conversely, an excessively high patience value can needlessly prolong the training duration, resulting in the inefficient use of computing resources and time. Moreover, there is no definite procedure for deciding an optimal value for patience; however, it depends on numerous factors like dataset characteristics, speed, algorithm complexity, and the available computational or hardware resources. Determining the optimal value

demands a process of experimentation and fine-tuning. By default, the patience value of the YOLOv5 model is 100, which was therefore used in our experiments. The patience parameter can be enabled or disabled (set patience = 0) easily. Fig. 7 illustrates the result of the patience parameter on the baseline model (YOLOv5n, YOLOv5s, and YOLOv5m) and the combination of baseline and attention models. Amongst the baseline models, YOLOv5n was trained for maximum epochs (324), whereas YOLOv5s was trained for 217 epochs, and YOLOv5m for 195 epochs. The fact that YOLOv5n trained for the maximum number of epochs among the baseline models suggests that it benefited from additional training time, potentially leading to better learning of complex features.



**Figure 7:** Results of the early stopping condition: baseline and proposed model

In the case of attention models, YOLOv5n + CBAM was trained for 492 epochs (highest), followed by YOLOv5n + ECA (446). However, the training of YOLOv5n + CBAM (263), YOLOv5n + ECA (286), and YOLOv5s + CBAM (272) stopped early. Fig. 7 also demonstrates the epoch at which the best results of each model were achieved. We aim to emphasize the impact of the ‘patience’ parameter on training duration and model performance. Notably, the attention models, particularly YOLOv5n + CBAM and YOLOv5n + ECA, reached higher epoch counts, indicating their ability to learn more effectively with the added complexity of the attention mechanisms. Hence, a higher patience value allows models to explore their learning capacity more fully, which can lead to improved performance. However, the early stopping of some models suggests that not all configurations benefit equally from extended training periods, highlighting the importance of monitoring performance metrics throughout training. Moreover, different datasets may yield varying training dynamics, potentially influencing how effectively the models learn over extended epochs. Therefore, while the results are promising for the datasets used in this study, they may not generalize across all scenarios.

## 5 Conclusion

This article has presented an innovative attention-enhanced YOLOv5 network for landslide detection from satellite imagery. The novelty lies in the integration of two lightweight though powerful attention models, specifically, CBAM and ECA, into the backbone and neck of the YOLOv5 architecture, respectively. Notably, the YOLOv5n + CBAM configuration achieved the highest detection accuracy among the evaluated

models, attaining a mAP@0.5 of 77.8%, closely followed by the YOLOv5n + ECA variant, which yielded a mAP@0.5 of 77.0%. The YOLOv5n + CBAM configuration demonstrated notable improvements in detection accuracy, outperforming existing methods by effectively focusing on key landscape features that are crucial for precise landslide mapping. The incorporation of attention mechanisms significantly enhances the model's ability to differentiate between subtle landscape variations, thereby improving its performance under complex and challenging terrain conditions. While the current findings are promising, they are still preliminary; the model's true potential will be more convincingly demonstrated through comprehensive case studies and large-scale image analysis in future work. We aim to explore additional YOLO variants and attention models to further refine and expand the capabilities of this approach. Eventually, the goal is to develop robust, real-time systems for landslide detection that can be seamlessly integrated into disaster management frameworks, thereby supporting mitigation efforts and enhancing resilience to natural hazards. This study contributes to the field by advancing machine learning applications in environmental monitoring, offering a promising solution to the challenges of real-time landslide mapping.

**Acknowledgement:** The authors would like to thank the director of the Wadia Institute of Himalayan Geology, Dehradun, for his motivation and encouragement. This work's contribution number is WIHG/255.

**Funding Statement:** This work is supported by the Department of Science and Technology, Science and Engineering Research Board, New Delhi, India, under Grant No. EEQ/2022/000812. Also funded by the Centre of Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney.

**Author Contributions:** Conceptualization, data collection, implementation, validation, manuscript writing, and funding acquisition: Naveen Chandra; implementation, validation, and manuscript editing: Himadri Vaidya; manuscript editing: Suraj Sawant; manuscript review and editing: Shilpa Gite; manuscript editing and critical review: Biswajeet Pradhan; funding for publication: Biswajeet Pradhan. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The dataset used for this study is freely available at <https://github.com/Abbott-max/dataset> (accessed on 01 January 2025).

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

1. Casagli N, Intrieri E, Tofani V, Gigli G, Raspini F. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat Rev Earth Environ*. 2023;4(1):51–64. doi:10.1038/s43017-022-00373-x.
2. Xu Y, Ouyang C, Xu Q, Wang D, Zhao B, Luo Y. CAS landslide dataset: a large-scale and multisensor dataset for deep learning-based landslide detection. *Sci Data*. 2024;11(1):12. doi:10.1038/s41597-023-02847-z.
3. Ma Z, Mei G, Piccialli F. Machine learning for landslides prevention: a survey. *Neural Comput Appl*. 2021;33(17):10881–907. doi:10.36227/techrxiv.12546098.v1.
4. Ma Z, Mei G. Deep learning for geological hazards analysis: data, models, applications, and opportunities. *Earth Sci Rev*. 2021;223:103858. doi:10.1016/j.earscirev.2021.103858.
5. Han Z, Fang Z, Li Y, Fu B. A novel Dynahead-Yolo neural network for the detection of landslides with variable proportions using remote sensing images. *Front Earth Sci*. 2023;10:1077153. doi:10.3389/feart.2022.1077153.
6. Liu P, Wei Y, Wang Q, Xie J, Chen Y, Li Z, et al. A research on landslides automatic extraction model based on the improved mask R-CNN. *ISPRS Int J Geo Inf*. 2021;10(3):168. doi:10.3390/ijgi10030168.
7. Lu P, Stumpf A, Kerle N, Casagli N. Object-oriented change detection for landslide rapid mapping. *IEEE Geosci Remote Sens Lett*. 2011;8(4):701–5. doi:10.1109/LGRS.2010.2101045.

8. Blaschke T, Feizizadeh B, Hölbling D. Object-based image analysis and digital terrain analysis for locating landslides in the Urmia Lake basin. Iran IEEE J Sel Top Appl Earth Obs Remote Sens. 2014;7(12):4806–17. doi:10.1109/JSTARS.2014.2350036.
9. Dong Z, An S, Zhang J, Yu J, Li J, Xu D. L-UNet: a landslide extraction model using multi-scale feature fusion and attention mechanism. Remote Sens. 2022;14(11):2552. doi:10.3390/rs14112552.
10. Keyport RN, Oommen T, Martha TR, Sajinkumar KS, Gierke JS. A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. Int J Appl Earth Obs Geoinf. 2018;64(6):1–11. doi:10.1016/j.jag.2017.08.015.
11. Ma Y, Wu H, Wang L, Huang B, Ranjan R, Zomaya A, et al. Remote sensing big data computing: challenges and opportunities. Future Gener Comput Syst. 2015;51(1):47–60. doi:10.1016/j.future.2014.10.029.
12. Zhang X, Zhou YN, Luo J. Deep learning for processing and analysis of remote sensing big data: a technical review. Big Earth Data. 2022;6(4):527–60. doi:10.1080/20964471.2021.1964879.
13. Huang Y, Zhao L. Review on landslide susceptibility mapping using support vector machines. Catena. 2018;165(1–2):520–9. doi:10.1016/j.catena.2018.03.003.
14. Sharma N, Saharia M, Ramana GV. High resolution landslide susceptibility mapping using ensemble machine learning and geospatial big data. Catena. 2024;235(1):107653. doi:10.1016/j.catena.2023.107653.
15. Wang H, Zhang L, Yin K, Luo H, Li J. Landslide identification using machine learning. Geosci Front. 2021;12(1):351–64. doi:10.1016/j.gsf.2020.02.012.
16. Lombardo L, Mai PM. Presenting logistic regression-based landslide susceptibility results. Eng Geol. 2018;244(1):14–24. doi:10.1016/j.enggeo.2018.07.019.
17. Hu Q, Zhou Y, Wang S, Wang F, Wang H. Improving the accuracy of landslide detection in “off-site” area by machine learning model portability comparison: a case study of Jiuzhaigou earthquake, China. Remote Sens. 2019;11(21):2530. doi:10.3390/rs11212530.
18. Fang Z, Wang Y, Duan G, Peng L. Landslide susceptibility mapping using rotation forest ensemble technique with different decision trees in the Three Gorges Reservoir area, China. Remote Sens. 2021;13(2):238. doi:10.3390/rs13020238.
19. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. IEEE Trans Pattern Anal Mach Intell. 2022;44(7):3523–42. doi:10.1109/tpami.2021.3059968.
20. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. arXiv:1807.05511v2. 2018.
21. Zeng D, Liao M, Tavakolian M, Guo Y, Zhou B, Hu D, et al. Deep learning for scene classification: a survey. arXiv:2101.10531. 2021.
22. Jia J, Ye W. Deep learning for earthquake disaster assessment: objects, data, models, stages, challenges, and opportunities. Remote Sens. 2023;15(16):4098. doi:10.3390/rs15164098.
23. Bentivoglio R, Isufi E, Jonkman SN, Taormina R. Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrol Earth Syst Sci. 2022;26(16):4345–78. doi:10.5194/hess-26-4345-2022.
24. Mohan A, Singh AK, Kumar B, Dwivedi R. Review on remote sensing methods for landslide detection using machine and deep learning. Trans Emerging Tel Tech. 2021;32(7):e3998. doi:10.1002/ett.3998.
25. Cheng G, Wang Z, Huang C, Yang Y, Hu J, Yan X, et al. Advances in deep learning recognition of landslides based on remote sensing images. Remote Sens. 2024;16(10):1787. doi:10.3390/rs16101787.
26. Shi W, Zhang M, Ke H, Fang X, Zhan Z, Chen S. Landslide recognition by deep convolutional neural network and change detection. IEEE Trans Geosci Remote Sens. 2021;59(6):4654–72. doi:10.1109/TGRS.2020.3015826.
27. Tehrani FS, Calvellido M, Liu Z, Zhang L, Lacasse S. Machine learning and landslide studies: recent advances and applications. Nat Hazards. 2022;114(2):1197–245. doi:10.1007/s11069-022-05423-7.
28. Akosah S, Gratchev I, Kim DH, Ohn SY. Application of artificial intelligence and remote sensing for landslide detection and prediction: systematic review. Remote Sens. 2024;16(16):2947. doi:10.3390/rs16162947.
29. Bragagnolo L, Rezende LR, da Silva RV, Grzybowski JMV. Convolutional neural networks applied to semantic segmentation of landslide scars. Catena. 2021;201(1):105189. doi:10.1016/j.catena.2021.105189.

30. Ghorbanzadeh O, Crivellari A, Ghamisi P, Shahabi H, Blaschke T. A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan). *Sci Rep.* 2021;11(1):14629. doi:10.1038/s41598-021-94190-9.
31. Ghorbanzadeh O, Xu Y, Ghamisi P, Kopp M, Kreil D. Landslide4Sense: reference benchmark data and deep learning models for landslide detection. *arXiv:2206.00515.* 2022.
32. Meena SR, Nava L, Bhuyan K, Puliero S, Soares LP, Dias HC, et al. HR-GLDD: a globally distributed dataset using generalized DL for rapid landslide mapping on HR satellite imagery. *Earth Syst Sci Data.* 2023;15(7):3283–98. doi:10.5194/essd-2022-350.
33. Devara M, Maurya VK, Dwivedi R. Landslide extraction using a novel empirical method and binary semantic segmentation U-NET framework using sentinel-2 imagery. *Remote Sens Lett.* 2024;15(3):326–38. doi:10.1080/2150704x.2024.2320178.
34. Liu T, Chen T, Niu R, Plaza A. Landslide detection mapping employing CNN, ResNet, and DenseNet in the Three Gorges Reservoir, China. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2021;14:11417–28. doi:10.1109/JSTARS.2021.3117975.
35. Ullo SL, Mohan A, Sebastianelli A, Ahamed SE, Kumar B, Dwivedi R, et al. A new mask R-CNN-based method for improved landslide detection. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2021;14:3799–810. doi:10.1109/JSTARS.2021.3064981.
36. Fu R, He J, Liu G, Li W, Mao J, He M, et al. Fast seismic landslide detection based on improved mask R-CNN. *Remote Sens.* 2022;14(16):3928. doi:10.3390/rs14163928.
37. Liu X, Xu L, Zhang J. Landslide detection with Mask R-CNN using complex background enhancement based on multi-scale samples. *Geomat Nat Hazards Risk.* 2024;15(1):2300823. doi:10.1080/19475705.2023.2300823.
38. Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl.* 2023;82(6):9243–75. doi:10.1007/s11042-022-13644-y.
39. Ju Y, Xu Q, Jin S, Li W, Su Y, Dong X, et al. Loess landslide detection using object detection algorithms in northwest China. *Remote Sens.* 2022;14(5):1182. doi:10.3390/rs14051182.
40. Pang D, Liu G, He J, Li W, Fu R. Automatic remote sensing identification of co-seismic landslides using deep learning methods. *Forests.* 2022;13(8):1213. doi:10.3390/f13081213.
41. Li B, Li J. Methods for landslide detection based on lightweight YOLOv4 convolutional neural network. *Earth Sci Inform.* 2022;15(2):765–75. doi:10.1007/s12145-022-00764-0.
42. Mo P, Li D, Liu M, Jia J, Chen X. A lightweight and partitioned CNN algorithm for multi-landslide detection in remote sensing images. *Appl Sci.* 2023;13(15):8583. doi:10.3390/app13158583.
43. Chandra N, Vaidya H. Automated detection of landslide events from multi-source remote sensing imagery: performance evaluation and analysis of YOLO algorithms. *J Earth Syst Sci.* 2024;133(3):127. doi:10.1007/s12040-024-02327-x.
44. Liu Q, Wu T, Deng Y, Liu Z. SE-YOLOv7 landslide detection algorithm based on attention mechanism and improved loss function. *Land.* 2023;12(8):1522. doi:10.3390/land12081522.
45. Mao Y, Niu R, Li B, Li J. Potential landslide identification based on improved YOLOv8 and InSAR phase-gradient stacking. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2024;17(5):10367–76. doi:10.1109/JSTARS.2024.3399788.
46. Hou H, Chen M, Tie Y, Li W. A universal landslide detection method in optical remote sensing images based on improved YOLOX. *Remote Sens.* 2022;14(19):4939. doi:10.3390/rs14194939.
47. Ji S, Yu D, Shen C, Li W, Xu Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides.* 2020;17(6):1337–52. doi:10.1007/s10346-020-01353-2.
48. Wang H, Liu J, Zeng S, Xiao K, Yang D, Yao G, et al. A novel landslide identification method for multi-scale and complex background region based on multi-model fusion: YOLO + U-Net. *Landslides.* 2024;21(4):901–17. doi:10.1007/s10346-023-02184-7.
49. Yang Y, Miao Z, Zhang H, Wang B, Wu L. Lightweight attention-guided YOLO with level set layer for landslide detection from optical satellite images. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2024;17:3543–59. doi:10.1109/JSTARS.2024.3351277.



50. Liu Q, Wu TT, Deng YH, Liu ZH. Intelligent identification of landslides in loess areas based on the improved YOLO algorithm: a case study of loess landslides in Baoji City. *J Mt Sci.* 2023;20(11):3343–59. doi:10.1007/s11629-023-8128-0.
51. Cheng L, Li J, Duan P, Wang M. A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides.* 2021;18(8):2751–65. doi:10.1007/s10346-021-01694-6.
52. Jia L, Leng X, Wang X, Nie M. Recognizing landslides in remote sensing images based on enhancement of information in digital elevation models. *Remote Sens Lett.* 2024;15(3):224–32. doi:10.1080/2150704x.2024.2313611.
53. Song Y, Guo J, Wu G, Ma F, Li F. Automatic recognition of landslides based on YOLOv7 and attention mechanism. *J Mt Sci.* 2024;21(8):2681–95. doi:10.1007/s11629-024-8669-x.
54. Zhang W, Liu Z, Zhou S, Qi W, Wu X, Zhang T, et al. LS-YOLO: a novel model for detecting multiscale landslides with remote sensing images. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2024;17:4952–65. doi:10.1109/jstars.2024.3363160.
55. Wang L, Lei H, Jian W, Wang W, Wang H, Wei N. Enhancing landslide detection: a novel LA-YOLO model for rainfall-induced shallow landslides. *IEEE Geosci Remote Sens Lett.* 2025;22:6004905. doi:10.1109/LGRS.2025.3541867.
56. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27–30; Las Vegas, NV, USA. p. 779–88. doi:10.1109/CVPR.2016.91.
57. Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach Learn Knowl Extr.* 2023;5(4):1680–716. doi:10.3390/make5040083.
58. Wang CY, Mark Liao HY, Wu YH, Chen PY, Hsieh JW, Yeh IH. CSPNet: a new backbone that can enhance learning capability of CNN. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; 2020 Jun 14–19; Seattle, WA, USA. p. 1571–80. doi:10.1109/cvprw50498.2020.00203.
59. Liu H, Sun F, Gu J, Deng L. SF-YOLOv5: a lightweight small object detection algorithm based on improved feature fusion mode. *Sensors.* 2022;22(15):5817. doi:10.3390/s22155817.
60. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 8759–68. doi:10.1109/CVPR.2018.00913.
61. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21–26; Honolulu, HI, USA. p. 936–44. doi:10.1109/CVPR.2017.106.
62. Mumuni A, Mumuni F. Data augmentation: a comprehensive survey of modern approaches. *Array.* 2022;16(6):100258. doi:10.1016/j.array.2022.100258.
63. Wang X, Song J. ICIOU: improved loss based on complete intersection over union for bounding box regression. *IEEE Access.* 2021;9:105686–95. doi:10.1109/access.2021.3100414.
64. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. *arXiv:1807.06521.* 2018.
65. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-net: efficient channel attention for deep convolutional neural networks. *arXiv:1910.03151.* 2019.
66. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 7132–41.
67. Zhao L, Liu J, Ren Y, Lin C, Liu J, Abbas Z, et al. YOLOv8-QR: an improved YOLOv8 model via attention mechanism for object detection of QR code defects. *Comput Electr Eng.* 2024;118(3):109376. doi:10.1016/j.compeleceng.2024.109376.
68. Liang F, Zhao L, Ren Y, Wang S, To S, Abbas Z, et al. LAD-Net: a lightweight welding defect surface non-destructive detection algorithm based on the attention mechanism. *Comput Ind.* 2024;161(7):104109. doi:10.1016/j.compind.2024.104109.

69. Wang T, Liu M, Zhang H, Jiang X, Huang Y, Jiang X. Landslide detection based on improved YOLOv5 and satellite images. In: 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI); 2021 Aug 20–22; Yibin, China. p. 367–71. doi:10.1109/prai53619.2021.9551067.
70. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions. arXiv:2112.05561. 2021.