

Article

From Landslide Detection to Multi-Source LLM-Based Reporting: A Complete Framework for Rapid Assessment of Post-Disaster Scenarios

Mohammed Alruqimi ¹, Abdelkader Riche ², Pierluigi Confuorto ^{2,*}, Mawloud Guermoui ¹,
Silvia Bianchini ² and Farid Melgani ¹

¹ Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38122 Trento, Italy; mohammed.alruqimi@unitn.it (M.A.); mawloud.guermoui@unitn.it (M.G.); farid.melgani@unitn.it (F.M.)

² Department of Earth Sciences, University of Florence, Via La Pira 4, 50121 Florence, Italy; abdelkader.riche@unifi.it (A.R.); silvia.bianchini@unifi.it (S.B.)

* Correspondence: pierluigi.confuorto@unifi.it

Highlights

What are the main findings?

- An end-to-end framework is proposed for landslide detection and reporting.
- Applied to a large real-world landslide event in Italy, the framework generates expert-style reports.

What are the implications of the main findings?

- The approach can support rapid post-event landslide screening and prioritization, reducing the time and effort needed for fully manual mapping and reporting.
- The framework offers a transferable blueprint for using remote sensing and AI to produce preliminary decision-support information for rapid post-disaster landslide assessment.

Abstract

Timely landslide detection and rapid qualitative assessment are fundamental to effective warning systems, hazard management, and risk mitigation. Yet, current practices that rely on on-site surveys and manual expert assessment remain risky, costly, and time-consuming. These limitations result in substantial delays between the event and the availability of actionable information. This study proposes a hybrid, multi-model framework that fuses RGB remote-sensing imagery with geospatial layers to enable timely landslide detection and actionable reporting. The pipeline couples an enhanced SegFormer (denoted as SDF-SegFormer-B2) model for landslide localization, a feature extraction technique for per-slide geo-attribute computation, and a lightweight instruction-tuned LLM (Mistral-7B-Instruct-v0.3) for structured, expert-style reporting. Although a few previous studies have explored landslide captioning, to our knowledge this is the first framework designed to generate structured technical reports enriched with terrain-context interpretation and qualitative intervention-priority indicators. Experiments use 26,758 georeferenced RGB tiles (64 × 64) with 3 m of spatial resolution from PlanetScope satellite imagery over Emilia-Romagna, Italy, with 68,592 annotated landslide boxes collected after the May 2023 rainfall events (~200 mm in 48 h on 1–3 May; 200–250 mm in 48 h on 16–17 May). The proposed SDF-SegFormer-B2 segmentation model achieved a precision of 85.54%, recall of 72.31%, and an F1-score of 78.39% on the unseen test dataset. To evaluate the quality of the generated landslide reports, 100 images were selected for domain-expert assessment. Among these, 58% of the reports were rated as “Very Good,” 30% as “Good,” 8% as “Acceptable,” and 4% as



Academic Editors: Fumio Yamazaki and Dongping Ming

Received: 5 March 2026

Revised: 27 May 2026

Accepted: 29 May 2026

Published: 2 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

“Poor.” When considering only reports with complete and accurate inputs, 81.48% were rated *“Very Good,”* and 96.30% were rated either *“Good”* or *“Very Good.”* By integrating complementary models and modalities, the proposed approach automates localization-to-reporting and enables the generation of terrain-aware landslide summaries that may support preliminary decision-making and rapid post-disaster screening.

Keywords: landslide reporting; landslide segmentation; landslide analysis; remote sensing images captioning; SegFormer; LLM

1. Introduction

Landslides are widespread geomorphic processes that threaten lives, livelihoods, and critical infrastructure, especially in mountainous settings and tectonically active regions. Their rapid onset and destructive power make timely detection and analysis indispensable for risk reduction, early warning, and efficient post-event response. Traditional field-based methods, while accurate, are time-consuming, costly, and often infeasible in remote or hazardous areas [1,2]. These limitations have motivated a shift toward remote sensing and artificial intelligence (AI), which together enable large-scale, data-driven approaches for landslide detection, segmentation, and susceptibility modeling [1,3].

Substantial progress has been made in automating landslide detection through deep learning. Object-detection architectures such as YOLO variants offer real-time localization with reduced computational cost [4–6], while semantic segmentation models—including U-Net, DeepLab, and SegFormer—enable pixel-level delineation of landslide footprints [2,7–12]. In parallel, studies have shown that integrating spectral imagery with ancillary geospatial attributes consistently improves both detection accuracy and model interpretability compared with RGB-only approaches [1,13–16].

Despite this progress, existing studies focus on detection or susceptibility mapping, leaving a critical gap between raw model outputs and the actionable information required for hazard management. Current post-event workflows still depend heavily on expert effort to interpret fragmented data and compile assessment reports. Meanwhile, recent advances in vision-language model (VLM) work for remote sensing largely emphasized captioning and related scene-description tasks [17,18]. Although effective for image-level interpretation, many of these models still lack explicit geospatial reasoning based on terrain and environmental attributes, which is essential for decision-oriented geohazard assessment. This limitation is important because landslide occurrence and severity are controlled by interacting topographic, geological, and material factors [19]. Therefore, credible post-event landslide reporting should combine image evidence with geospatial context, which can reduce errors and hallucinations and improve practical usefulness.

To address the abovementioned gaps, this study presents a framework for automated landslide detection and reporting to support actionable geohazard intelligence. The system follows a multi-stage pipeline: (i) detect and localize landslides in remote-sensing imagery; (ii) extract geo-features for affected areas; and (iii) generate structured, expert-style reports. The resulting outputs include geospatial characterization, analytical interpretation, and concise priority intervention recommendations. For landslide localization, a SegFormer architecture was extended by replacing the original loss function with a combined Focal–Dice loss and integrating a slope channel through a lightweight auxiliary encoder. To enrich each detected region with physical context, threshold-based feature extraction techniques are applied to ingest ancillary geospatial layers—Digital Elevation Model (DEM), slope, Normalized Difference Vegetation Index (NDVI), and lithology—and infer key geo-

attributes. These attributes, together with the detected regions, are then provided to a Large Language Model (LLM) that produces comprehensive landslide reports.

The technical contribution of this work lies in three key areas:

First, a geospatially grounded LLM architecture: The proposed framework explicitly conditions the language model on physically meaningful geo-attributes. This grounding strategy addresses a fundamental challenge in geoscientific LLM applications, transforming the output from generic, hallucination-prone image captions into rigorous, expert-level reports that are both physically consistent and geospatially reliable. Second, enhancing pixel-level localization accuracy with the introduced SDF-SegFormer-B2 model. Third, end-to-end automation for operational deployment: the framework introduces a streamlined pipeline from satellite imagery acquisition to preliminary assessment report generation for post-disaster response support.

This paper is structured as follows: Section 2 describes the dataset and the proposed framework with the experimental setup; Section 3 presents the results and evaluation; Section 4 provides discussion and outlines limitations; and Section 5 concludes the paper and highlights directions for future work.

2. Materials and Methods

2.1. Study Area and Dataset

The dataset used in this study comprises 26,758 georeferenced RGB tiles sourced from PlanetScope (<https://www.planet.com/explorer>, accessed on 20 January 2025) in GeoTIFF format. Each tile is 64×64 pixels at 3 m spatial resolution. The imagery covers the Emilia–Romagna region of Italy. Tile selection was guided by the availability and quality of post-event satellite imagery. Due to cloud coverage in some areas immediately after the event of 1–17 May 2023, imagery was selected within a temporal window of up to 30 days following the event. Priority was given to cloud-free images acquired as close as possible to the event date, while in other cases, the most suitable imagery within this timeframe was used. The dataset was divided into 19,324 tiles for training, 4832 for validation, and 2602 for testing. Furthermore, the dataset exhibits class imbalance between landslide and non-landslide pixels. Landslide pixels represent approximately 7.75% of the training set and 7.60% of the validation set, while in the test set, they account for about 5.37%, with non-landslide pixels exceeding 90% across all subsets. In May 2023, two extreme rainfall events hit the region: the first (1–3 May) delivered ~200 mm in 48 h; the second (16–17 May) brought 200–250 mm over the same duration, affecting overlapping zones [20–22]. Each event corresponds to a 100–300-year return period; together, they represent a 500-year event. Image selection followed a 30-day post-event window to mitigate cloud cover: we prioritized the earliest cloud-free scenes and otherwise used the best available imagery within that window. In addition to the RGB data, each tile is supported with four auxiliary geospatial layers: digital elevation model (DEM), slope, normalized difference vegetation index (NDVI), and lithology.

Figure 1 situates the study area within Italy, illustrating the terrain context alongside the susceptibility map by [23], which delineates extensive high to very high susceptibility belts along the northern Apennines. Geologically, the area spans the northern Apennines and includes >600 formations from Jurassic to Miocene, dominated by sedimentary successions (flysch, marls, clays, and tectonic mélanges) [24]. The western–central sector (Piacenza–Bologna) is mainly Ligurian/Sub-Ligurian, with clay-rich, low-strength formations and block-in-matrix mélanges (the “Argille Scagliose”) that are especially failure-prone; by contrast, the eastern Apennines and ridge zones are underlain by more competent Tuscan–Umbrian sandstone–marl sequences, which are comparatively more stable. Landslides are annotated with 68,592 bounding boxes, which we generated by convert-

ing polygons from the [20] landslide inventory shapefile (publicly available at Zenodo: <https://zenodo.org/records/13742643>, accessed on 28 May 2026) into bounding boxes. The statistical distribution of mapped landslides reveals a pronounced dominance of small-area events and a limited number of large failures. All image tiles in the dataset contain at least one or more landslides, ensuring full spatial relevance for detection and characterization tasks. As shown in Figure 2 (left), the area distribution is heavily concentrated toward smaller landslide sizes, with only a few instances representing large surface extents. Similarly, Figure 2 (right) demonstrates that the majority of image tiles contain between 1 and 5 landslide instances, with frequency rapidly declining as the number of occurrences increases. This behavior reflects the natural imbalance of landslide inventories, where minor slope failures are far more frequent than large-scale movements.

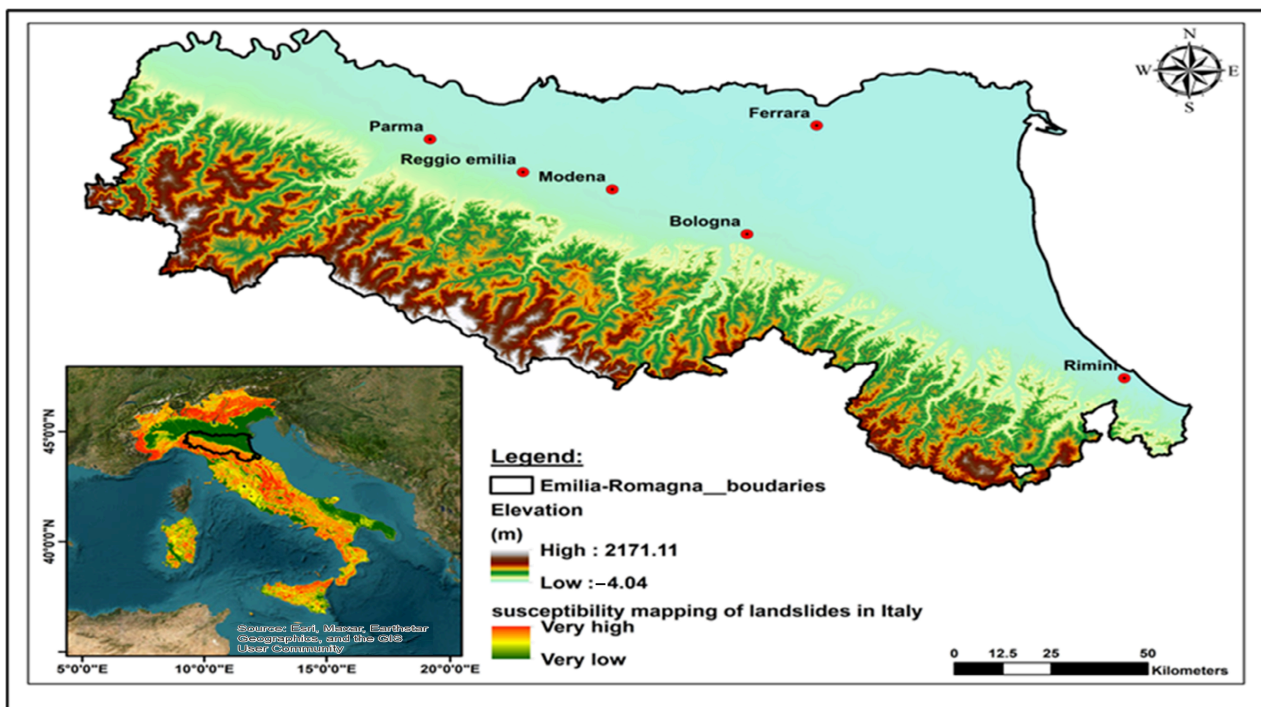


Figure 1. Study area location. The main panel shows the Emilia–Romagna administrative boundary over a DEM, the inset locates Emilia–Romagna within Italy, and displays the national landslide susceptibility map.

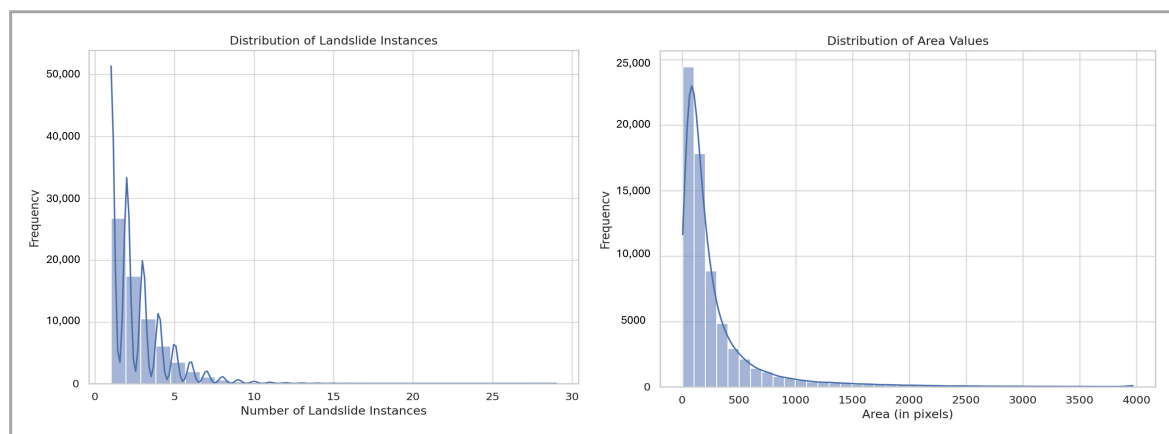


Figure 2. Distribution of landslide instances and area values.

2.2. Proposed Model Framework

The proposed framework (as illustrated in Figure 3) combines RGB imagery with geospatial layers to generate structured, domain-specific reports for landslide-affected areas. To this end, the framework follows a three-stage pipeline: (i) landslide identification and localization; (ii) per-slide geo-attribute extraction from auxiliary rasters; and (iii) report generation with a compact instruction-tuned LLM. This end-to-end design eliminates the need for region proposal steps, streamlining the pipeline and substantially improving both speed and efficiency. The three stages are described below.

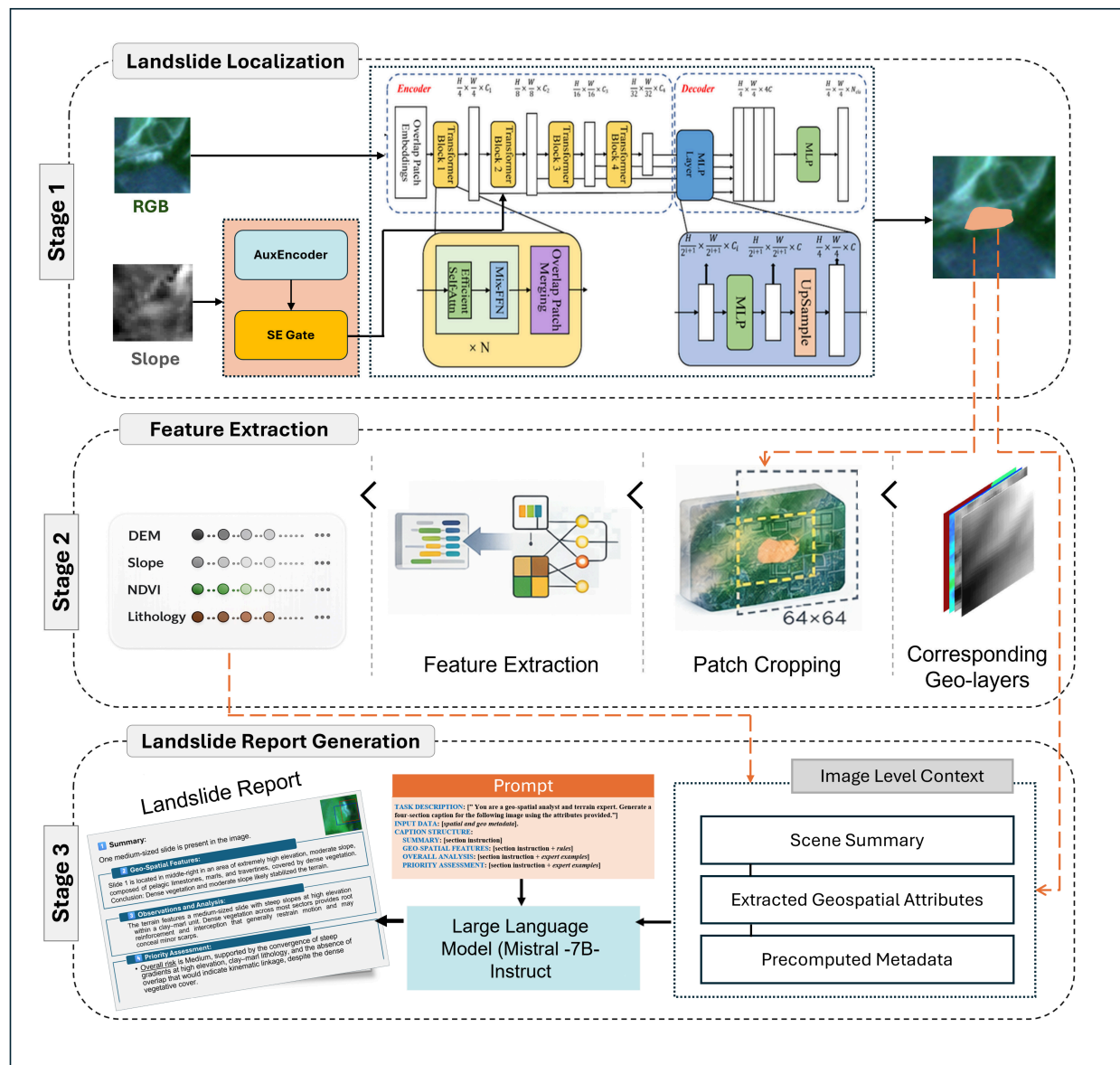


Figure 3. The proposed framework consisting of three stages: Stage 1—SDF-SegFormer-B2 landslide identification and localization; Stage 2—Feature extraction from auxiliary geo-layers; Stage 3—LLM-based landslide reporting.

2.2.1. Stage 1: Landslide Localization

The first stage of the proposed framework focuses on identifying and localizing landslides in remote sensing imagery. To achieve this, two distinct approaches were investigated: object detection using YOLOv11 (bounding-box localization) and semantic segmentation (pixel-wise delineation) via SegFormer. Multiple variants of SegFormer-based architectures

were evaluated. The SegFormer architecture [25] consists of a hierarchical encoder and a lightweight MLP decoder designed to balance global context with local detail.

The encoder is divided into four Transformer blocks (Blocks 1–4) that extract multi-scale features through three key modules: Overlap Patch Embedding (OPE) for preserving edge continuity, Efficient Multi-head Self-Attention (EMSA) for modeling long-range dependencies with low complexity, and a Mix Feed-Forward Network (Mix-FFN). An important innovation is the use of 3×3 depth-wise convolutions within the Mix-FFN to provide spatial awareness; this eliminates the need for fixed positional encodings, allowing the model to remain robust across various input resolutions. The decoder then takes the hierarchical feature maps F_i , where i in $\{1, 2, 3, 4\}$, upsamples them to a unified resolution of $\frac{H}{4} \times \frac{W}{4} \times C$, and concatenates them. This simple MLP-based approach fuses multi-level features to generate the final segmentation map, achieving high precision on complex boundaries while maintaining a highly efficient and streamlined design.

In this work, a baseline configuration was established by training a compact SegFormer model on RGB imagery alone using weighted cross-entropy. Building on this, we performed an ablation study. Specifically, we evaluated various SegFormer backbone capacities, tested a compound loss function (Focal + Dice), and explored the integration of slope as an auxiliary topographic band through an AuxEncoder. This lightweight auxiliary CNN encoder, composed of three stride-2 convolutional layers (~140 K parameters), maps the slope map to the Transformer Block-2 feature resolution. A Squeeze-and-Excitation gate then uses globally pooled auxiliary features to generate per-channel attention weights that modulate the Block-2 RGB hidden state before it enters the MLP decoder. This early-stage injection preserves spatial granularity while providing sufficient semantic context for terrain-aware feature selection. The loss function (Focal + Dice) is mathematically expressed in Equations (1)–(3) as follows:

$$\mathcal{L} = 0.5\mathcal{L}_{\text{focal}} + 0.5\mathcal{L}_{\text{dice}} \quad (1)$$

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N} \sum_i \alpha_t (1 - p_t)^\gamma \log(p_t) \quad (2)$$

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2 \sum_i p_i^{(1)} y_i + 1}{\sum_i p_i^{(1)} + \sum_i y_i + 1} \quad (3)$$

where N is the number of pixels, p_t is the predicted probability for the ground-truth class, α_t in $\{0.75, 0.25\}$ is the class-balancing weight for foreground and background, respectively. γ is the focusing parameter that down-weights easy examples, $p_i^{(1)}$ is the predicted foreground probability at pixel i , and y_i in $\{0, 1\}$ is the ground-truth label. The auxiliary features are fused into the encoder, where a band (e.g., slope) branch generates a channel-wise attention vector that recalibrates the Block-2 RGB features, as expressed by Equations (4) and (5).

$$\alpha = \sigma(W_2 \delta(W_1 \cdot \text{GAP}(f_s))) \quad (4)$$

$$h_2^* = h_2 \odot \alpha \quad (5)$$

where GAP denotes global average pooling, δ is ReLU, σ is the sigmoid activation, and (f_s) is the slope feature map at Block-2 resolution. W_1 and W_2 are the learnable weight matrices of two fully connected layers inside the SE gate. The resulting h_2^* replaces h_2 before propagating to the subsequent encoder stages. Additionally, an object-detection strategy was evaluated based on YOLOv11 [26] and RT-DETRv2 [27].

2.2.2. Stage 2: Geo-Attribute Extraction

In this stage, auxiliary geospatial layers (DEM, slope, NDVI, and lithology) are retrieved for each landslide instance detected in Stage 1. These layers capture essential terrain, vegetation, and lithological characteristics that influence landslide occurrence. From these datasets, geospatially referenced features were extracted corresponding to the location of each detected landslide. The extracted values are then processed through a threshold-ranging procedure to classify continuous variables (e.g., elevation, slope, NDVI) and group lithological units into meaningful categories. This process generates a per-landslide attribute set that integrates spatial, topographic, and environmental descriptors with the detection metadata obtained from the first stage. Such integration provides a robust contextual foundation for the LLMs to perform in-depth technical analysis and interpretative assessment. The integrated geo-attribute layers and the procedure employed for the threshold ranging are described as follows.

- **DEM:** A 5 m DEM from the Emilia–Romagna Geoportal (<https://geoportale.regione.emilia-romagna.it/catalogo/dati-cartografici/altimetria/layer-2>, accessed on 28 May 2026) was used and resampled to 3 m to match PlanetScope tiles. Elevation reflects relief and gravitational potential, modulating drainage and failure patterns. In this study, the DEM was classified as moderate–low elevation (41.34–198.50 m), high elevation (198.50–355.66 m), and extremely high elevation (355.66–512.82 m). The thresholds were defined by dividing the observed elevation range (~41–513 m) into three equal intervals to ensure balanced representation across terrain classes. This equal-interval approach has been widely used in susceptibility and geomorphological studies when no universally accepted elevation thresholds exist, as it allows meaningful comparison between lowland, mid-altitude, and upland environments [28].
- **Slope:** It was computed from the 5 m DEM, which had been resampled to 3 m, using the (planar gradient in degrees). Because steepness increases driving stresses while moderate gradients can store colluvium that mobilizes during intense rainfall, the slope was classified as moderate (0.00–15.00°) and steep (15.00–45.00°). The threshold of 15° was selected based on practical applications and land suitability classifications reported in previous studies, where gentle to moderate slopes (<15°) are often considered stable for agriculture and settlement, while steeper slopes (>15°) are more prone to instability and erosion [28]. This two-class division also provides a parsimonious scheme that captures the primary contrast between low-gradient depositional zones and higher-gradient unstable slopes within the Emilia–Romagna study area.
- **NDVI:** Was derived from PlanetScope NIR and red bands using the standard formula $NDVI = (NIR - Red) / (NIR + Red)$ [29]. To reflect canopy density and related root reinforcement, NDVI was classified as Low vegetation (NDVI < 0.4) and Dense vegetation (NDVI > 0.4). The threshold of 0.4 was chosen because lower values typically correspond to sparse or stressed vegetation, whereas higher values indicate closed canopy cover and denser biomass [30].
- **Lithology:** Was rasterized from the National Lithological Map of Italy (1:500,000), available through the ISPRA geological web viewer: <https://sgi2.isprambiente.it/viewersgi2>, accessed on 28 May 2026) to the 3 m grid. Because material strength, fabric, permeability, and swelling behavior govern failure style and frequency, lithology was grouped into 5 classes: (1) clays and marls; (2) alluvium, debris, and glacial deposits; (3) pelagic limestones, marls, and travertines; (4) sands, sandstones, and conglomerates; and (5) gypsum, anhydrite, and evaporitic salts.

Figure 4 provides representative samples across the RGB and all GEO-layers (DEM, slope, NDVI, and lithology), and Table 1 summarizes the class thresholds used for each layer.

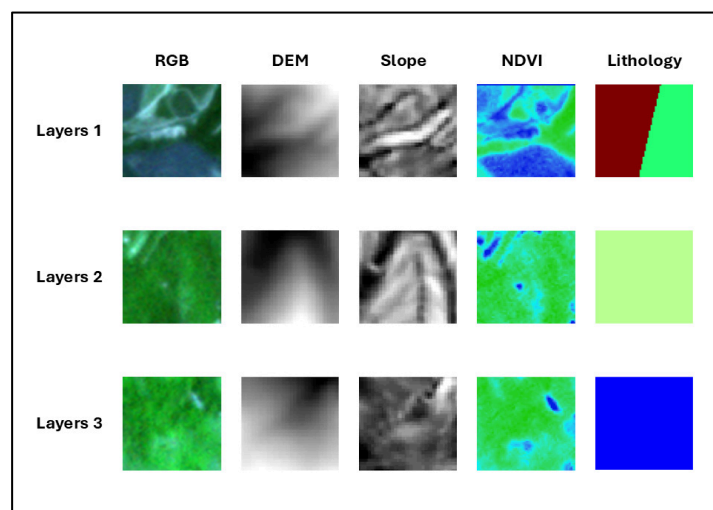


Figure 4. RGB images and corresponding auxiliary geo-layers.

Table 1. Classes and thresholds for geo-layers.

Layer	Class	Range
DEM	Moderate–low elevation	41.34–198.50 m
	High elevation	198.50–355.66 m
	Extremely high elevation	355.66–512.82 m
Slope	Moderate slope	0.00–15.00°
	Steep slope	15.00–45.00°
NDVI	Low vegetation	<0.4
	Dense vegetation	>0.4
Lithology	Class 1	Clays and marls
	Class 2	Alluvium, debris, and glacial deposits
	Class 3	Pelagic limestones, marls, and travertines
	Class 4	Sands, sandstones, and conglomerates
	Class 5	Gypsum, anhydrite, and evaporitic salts

2.2.3. Stage 3: Report Generation with Mistral

For the report generation stage, we employed Mistral-7B-Instruct-v0.3, a state-of-the-art open-source instruction-tuned large language model developed by Mistral AI [31]. This model builds upon the Mistral-7B architecture, which utilizes a dense transformer design optimized for efficiency and high-quality text generation. Mistral-7B-Instruct-v0.3 is specifically fine-tuned for instruction following and reasoning tasks, offering a strong balance between efficiency, fluency, and contextual understanding despite its relatively compact size compared to larger LLMs. To guide the landslide reporting process, we designed structured prompts enriched with metadata derived from the previous two stages (Stages 1 and 2), as shown in Figure 5. The metadata comprises the geographic coordinates of detected landslides identified in Stage 1, which are further utilized to compute spatial descriptors such as the area of each landslide and the relative distances among multiple instances within a scene. Additionally, geo-attributes generated in Stage 2 are incorporated for each detected landslide to provide detailed contextual information. To enhance the model's capability in this domain-specific task, a few-shot learning strategy was employed. This approach embeds a set of expert-curated exemplars within the prompting context, enabling the model to infer task-specific patterns and generate analytically consistent, domain-relevant outputs. Specifically, six exemplars were hand-crafted from scratch by a domain expert (geologist), independently of the training and test sets, and chosen to cover

a diverse range of terrain and lithological regimes—including clastic, carbonate, evaporitic, and clay–marl settings under varying slope, elevation, and vegetation conditions.

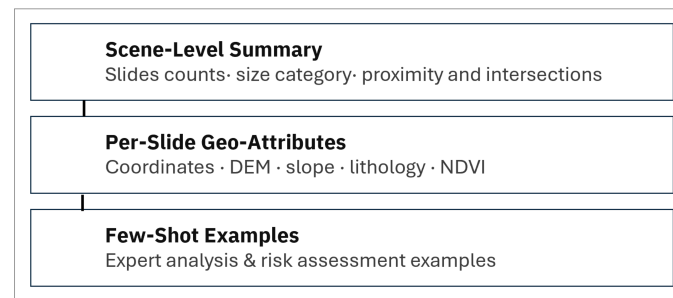


Figure 5. LLM prompt metadata input.

2.3. Experimental Setup

All experiments were conducted on a Linux system equipped with a GPU GeForce RTX 3090. As detailed above, the implementation followed a three-stage pipeline: (1) landslide localization with SegFormer, (2) geo-attributes extraction and thresholding, and (3) report generation with Mistral-7B-Instruct-v0.3.

SegFormer has been adopted as the foundation of our segmentation framework. We systematically ablate three design axes: (i) the backbone capacity (MiT-B0, ~3.7 M parameters, vs. MiT-B2, ~25.4 M parameters), (ii) the training loss function, and (iii) the integration of auxiliary topographic information. Table 2 summarizes all evaluated architectures, including the SegFormer variants and the YOLOv11 and RT-DETRv2 object-detection baselines. Both backbones were initialized from ImageNet-1k pretrained checkpoints (NVIDIA/mit-b0 and NVIDIA/mit-b2), while the AuxEncoder (~140 K parameters) was trained from scratch jointly with the main network. Training employs the AdamW optimizer ($\text{lr} = 6 \times 10^{-5}$, weight decay = 0.01) with a linear warm-up over the first five epochs followed by cosine annealing. Mixed-precision training (AMP) is used for all MiT-B2 variants to reduce memory footprint and accelerate convergence.

Table 2. Evaluated model architecture.

Model	Backbone	Loss Function	Input	Params
SegFormer-B0	MiT-B0	Weighted Cross-Entropy	RGB	~3.7 M
FD-SegFormer-B0	MiT-B0	Focal + Dice	RGB	~3.7 M
FD-SegFormer-B2	MiT-B2	Focal + Dice	RGB	~25.4 M
SDF-SegFormer-B2	MiT-B2	Focal + Dice	RGB + Slope	~25.5 M
YOLOv11	YOLOv11n	BCE + CIoU + DFL	RGB	~2.6 M
RT-DETRv2	ResNet-101vd	VFL + L1 + GIoU	RGB	~77.4 M

YOLOv11 and RT-DETRv2 were additionally evaluated as object-detection benchmarks, while the proposed pipeline itself relies on SegFormer. Since the two detectors differ substantially in architecture, each was fine-tuned with the training recipe recommended by its official repository rather than with artificially matched settings. YOLOv11 was initialized from the COCO-pretrained yolo11n checkpoint (Ultralytics) [26] using its default optimizer, augmentation stack, and loss formulation: BCE for classification, CIoU for bounding-box regression, and DFL for distribution focal loss. RT-DETRv2 was initialized from the COCO-pretrained rtdetr_v2_r101vd checkpoint (ResNet-101vd backbone, Hugging Face release PekingU/rtdetr_v2_r101vd) [27] and fine-tuned using a loss formulation comprising Vari-focal Loss (VFL), L1, and GIoU for bounding-box regression. Both baselines share the same train/validation/test split as SegFormer to guarantee a fair comparison at the data level. In

terms of computational performance, the localization stage was efficient on an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM under single-image end-to-end inference, using a batch size of 1. The full 2,602-tile test set was processed in approximately 33 s for YOLOv11 (~11 ms/tile), 37 s for SDF-SegFormer-B2 (~14 ms/tile), and 92 s for RT-DETRv2 (~29 ms/tile). The end-to-end runtime is dominated by the report generation stage, where Mistral required approximately 11 s per tile to generate a complete report. This results in an end-to-end processing time dominated by the LLM inference, highlighting a trade-off between report quality and latency in the current framework. For performance assessment, we conducted both quantitative and qualitative evaluations. Quantitative assessments were performed separately for the detection stage to measure the accuracy and effectiveness of this component independently. Subsequently, domain experts evaluated the generated reports for correctness, completeness, and domain-specific quality.

3. Results

3.1. Landslide Localisation

Table 3 summarizes the top-performing SegFormer configurations retained according to validation metrics, in addition to the results of YOLOv11 and RT-DETRv2. Out of 2602 images in the test dataset, containing a total of 8533 ground truth landslides, the model correctly detected 6170 instances (TP), with 1045 false positives (FP) and 2363 false negatives (FN). Precision P , recall R , and $F1$ -score are computed as $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$, and $F1 = \frac{2 \cdot (P \cdot R)}{P+R}$, yielding the values reported in Table 3 (85.54%, 72.31%, and 78.39%, respectively). Visual comparisons between the ground truth and the evaluated models' predictions are presented in Figures 6 and 7. The most pronounced behavioral shift arises from replacing weighted cross-entropy with the combined Focal–Dice loss, accompanied by an updated training regime and linear warm-up scheduling. This is reflected most directly in the precision jump (e.g., 77.82% to 82.11% when moving to FD-SegFormer-B0), even though it comes with a recall trade-off (74.13% to 69.12%), suggesting the loss combination is making the model more conservative (fewer false positives but also missing some positives). Scaling up the backbone to MiT-B2 and retaining Focal–Dice further improves overall balance, with FD-SegFormer-B2 and especially SDF-SegFormer-B2 (RGB + Slope) producing the most stable masks and the best overall scores (highest $F1$ and IoU , with SDF-B2 leading at 78.39% $F1$ /63.92% IoU). Finally, YOLOv11 lags in overlap (IoU 52.78%), consistent with less accurate pixel-level localization compared with the segmentation models. However, for the object detection models, the results reveal a clear trade-off: RT-DETRv2 attains higher recall by identifying more landslide instances but introduces substantial false positives (low precision), whereas YOLOv11 achieves higher precision with fewer false detections at the expense of lower recall. RT-DETRv2 could be interpreted as a high-recall object-detection benchmark rather than as the best-performing operational model in the proposed pipeline. Its IoU of 59.37% indicates reasonable spatial overlap for matched detections, but the large number of unmatched false-positive boxes substantially reduces its precision to 30.79%.

Table 3. Evaluation of landslide localization performance on the test dataset.

Model	Precision	Recall	F1-Score	IoU
SDF-SegFormer-B2	85.54%	72.31%	78.39%	63.92%
FD-SegFormer-B2	83.46%	72.07%	77.35%	63.06%
FD-SegFormer-B0	82.11%	69.12%	75.03%	58.60%
SegFormer-B0	77.82%	74.13%	75.94%	61.12%
YOLOv11	78.26%	62.94%	69.76%	52.78%
RT-DETRv2	30.79%	75.92%	43.81%	59.37%

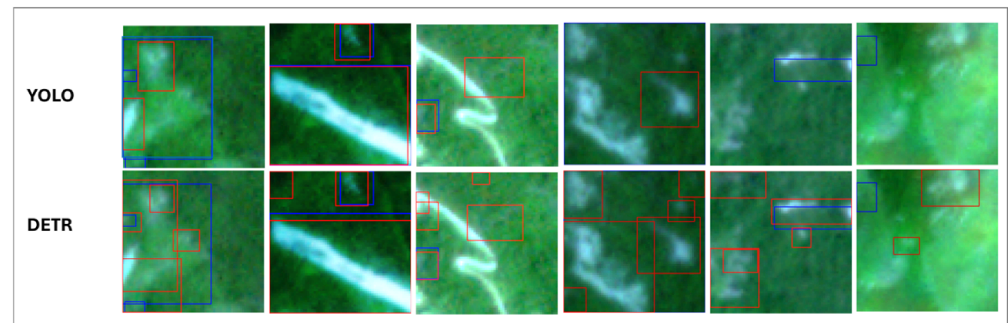


Figure 6. Visual comparison of landslide object detection results using YOLO and RT-DETRv2 (blue boxes are the GT and the red boxes are the predictions).

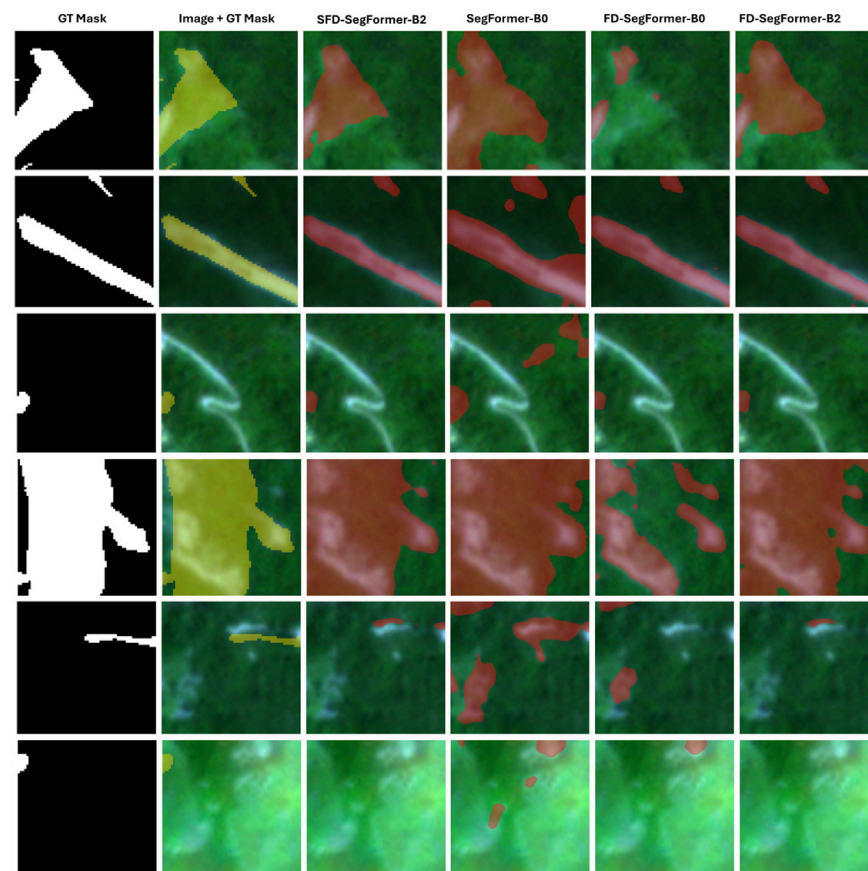


Figure 7. Visual comparison of landslide pixel-level segmentation across different SegFormer-based configurations.

On the other hand, false negatives are mainly observed in areas where satellite image acquisition occurred with a temporal delay following the landslide event, in some cases exceeding three weeks. In such situations, partial vegetation regrowth and natural surface recovery can obscure landslide signatures, reducing their visibility in RGB imagery and making detection more challenging. These observations highlight the influence of both spectral similarity and acquisition timing on model performance, particularly in complex or evolving post-event environments. To provide a comprehensive view of the model performance, an overall spatial representation of landslide detection results over the test area is presented in Figure 8. This visualization highlights the distribution of true positives, false positives, and false negatives, offering insights into the behavior of the model under varying landscape conditions. The analysis indicates that some false positive detections occur in regions with spectral characteristics similar to landslides, such as bare soil, exposed

rock, or recently disturbed agricultural surfaces, which exhibit reflectance patterns close to actual landslide areas.

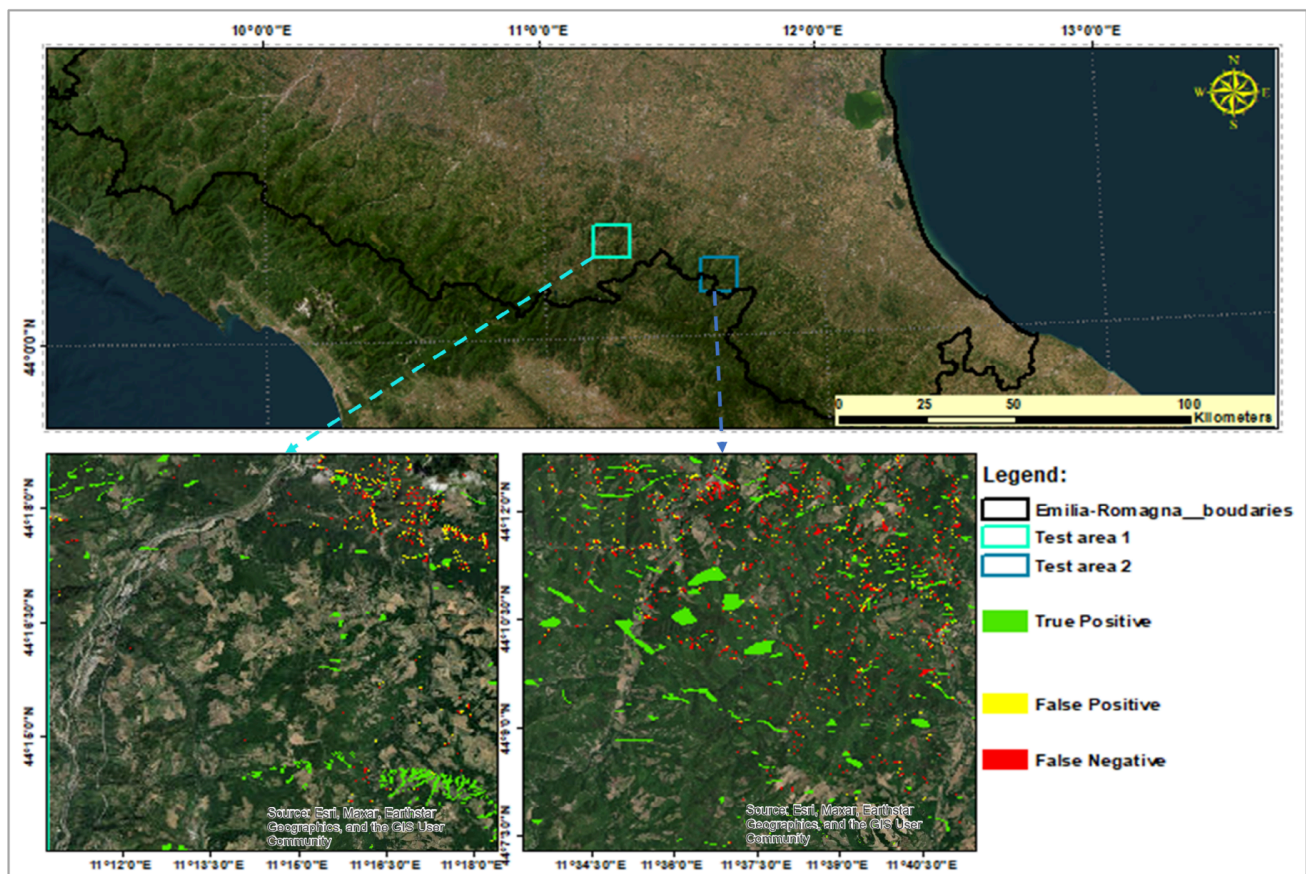


Figure 8. Overview of landslide detection performance across the test area, highlighting correct detections and misclassification patterns.

3.2. Report Generation

Structured terrain context reports represent the final output of the framework. For each input image, the model is instructed to generate a four-section structured report, as shown in Figure 9.

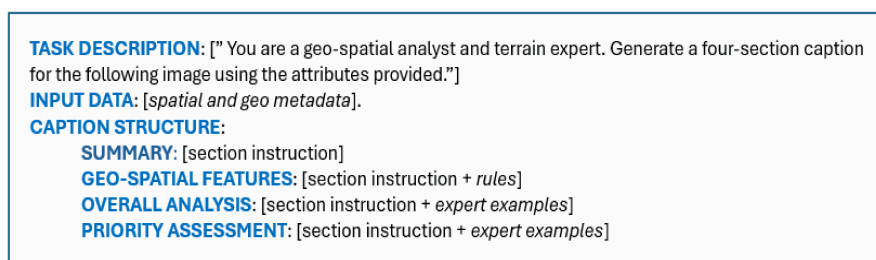


Figure 9. Pseudo-template of the LLM prompt for report generation.

(i) Summary: presents a statistical overview of the image, encompassing the number of landslides, their size classifications, and the presence of any overlapping occurrences. The size classification is defined as follows (in number of pixels, where each pixel represents an area of 3×3 m): “very large” (>3000), “large” (2001–3000), “medium” (501–2000), “small” (101–500), and “very small” (0–100).

(ii) Geospatial Features: provides a slide-by-slide characterization of DEM, slope, NDVI, and lithology, accompanied by a deterministic rule-based conclusion derived from

these features. For instance, sparse vegetation combined with steep slopes is interpreted as accelerating landslide activity, while dense vegetation on moderate slopes is associated with terrain stabilization. Similarly, clay-rich lithologies with vegetation cover suggest resistance to spread, whereas sandy substrates on slopes promote flow-like failures. These rule-based conclusions are provided as grounding inputs to the LLM to constrain its interpretation and avoid hallucination.

(iii) Overall Analysis: presents a comprehensive scene-level geological interpretation that synthesizes the detected landslides with their associated geo-attributes.

(iv) Priority Assessment: provides both overall and detailed evaluations of the landslide. The assessment includes three elements. First, it indicates the priority intervention level, categorized as low, medium, or high, with a justification synthesized from all detected landslides. These priorities were determined qualitatively based on terrain elevation, slope steepness, lithological strength, vegetation cover, and the spatial proximity of multiple failures, which together reflect the urgency of potential intervention. A qualitative overall priority indicator is inferred by the LLM based on the combination of these terrain attributes and the spatial configuration of detected slides (overlap, adjacency), guided by domain-expert-curated few-shot examples embedded in the prompt that enforce consistent reasoning patterns grounded in geomorphological principles. Second, it reports the conditional outlook in prose for three scenarios: short intense rainfall, prolonged wet season or snowmelt, and vegetation loss due to NDVI decline, harvesting, or wildfire. Each scenario is described within the text and assigned a likelihood level of low, medium, or high. Additionally, the report includes a probable velocity estimate classified according to the Cruden and Varnes [32] scale, which defines seven levels: extremely slow, very slow, slow, moderate, rapid, very rapid, and extremely rapid. The probable velocity is reported as complementary contextual information intended to support the interpretation of the detected landslide behavior. However, it is not directly incorporated into the priority assessment score, since no direct kinematic measurements or temporal displacement observations are available within the scope of the proposed rapid post-disaster framework. Consequently, the overall priority and the overall risk are derived exclusively from the extracted geo-attributes and surrounding spatial configuration. The reported velocity class should therefore be interpreted as a qualitative approximation of potential movement behavior rather than as a measured parameter used in the operational prioritization process, and it is not intended to replace physically based analyses or monitoring-driven evaluations when direct displacement or velocity measurements are available. To ensure a reliable evaluation, the reports were assessed qualitatively with the support of three domain experts. One expert conducted the evaluation, which was subsequently verified by the other two experts. The two additional experts independently reviewed the initial assessments to confirm their validity and consistency. In cases of uncertainty or disagreement, a consensus was reached through discussion, ensuring that the final evaluation reflects a validated agreement among all experts rather than a single subjective judgment. For this evaluation, 100 tiles were selected from the detection test set using guided stratified sampling. Each tile was labeled according to the Stage 1 localization outcome, including correct detections, false positives, false negatives, and mixed cases, and according to the Stage 2 geo-attribute classes, including DEM, slope, NDVI, and lithology. The sampling procedure considered the available classes within each auxiliary layer (e.g., DEM elevation classes, slope-gradient classes, NDVI vegetation classes, and lithological categories), and the final subset was selected to represent the observed combinations between these geo-attribute classes across the different tiles. The final subset was chosen to cover the main observed combinations of detection quality and terrain conditions while also preserving variation in the number and size of landslide instances, thereby preventing the qualitative assessment from being

biased toward a specific localization pattern or terrain regime in the study area. For each image, the Mistral model was prompted with geospatial metadata obtained and derived from Stage 1 and Stage 2. The reports were then evaluated manually across multiple dimensions, addressing the four report sections from different perspectives, as detailed in the following subsection.

3.2.1. Reports Evaluation Concept

To ensure a systematic and comprehensive assessment of the generated reports, a manual expert-based evaluation was performed following a structured evaluation rubric grounded in three complementary metrics: Score, Noisy Information, and Missing Information, as defined in Table 4. Each metric was independently applied to evaluate the quality of specific sections within the generated reports. The evaluation was limited to Section 1, Section 3, and Section 4, whereas Section 2 was excluded from the expert evaluation because it is structured metadata deterministically extracted from the source rasters using the fixed threshold rules defined in Table 1. Subsequently, a Global Score is derived to provide an integrated measure that reflects the overall quality and coherence of the report. For the scoring metric, the possible values are defined as follows: ‘*Very Good*’ when the caption is free of irrelevant information and provides strong domain-focused interpretation with coherent terrain-context analysis. ‘*Good*’ when only minor issues are present, such as a missing attribute that does not significantly affect quality, while the report still delivers solid domain-centered interpretation and useful contextual analysis. ‘*Acceptable*’ when issues arise due to the inclusion of irrelevant information, which slightly undermines clarity, but the report still offers meaningful contextual insights overall. ‘*Poor*’ applies to all other cases.

Table 4. Evaluation metrics of the generated reports.

Metric	Description	Possible Values
Score	Evaluates report quality in terms of reasoning, relevance, and depth from a domain-expert perspective.	Very good, good, acceptable, and poor
Noisy Info	False positive (FP): any factual mismatch between the report and the ground truth, i.e., any incorrect statement about the four ground-truth geo-attributes (DEM, slope, lithology, and NDVI) or about the ground-truth landslide instances (e.g., landslides reported where none exist, incorrect count, or mis-delineated boundaries). The flag is set to Yes if at least one such mismatch is found in the evaluated section.	Yes, No
Missing Info	False negative (FN): any omission of information that is present in the ground truth, i.e., any of the four ground-truth geo-attributes (DEM, slope, lithology, NDVI) or any ground-truth landslide instance that is not mentioned in the report. The flag is set to Yes if at least one such omission is found in the evaluated section.	Yes, No

The domain experts apply their own knowledge together with the available ground truth data, including slide bounding boxes and geospatial attributes, to ensure a robust and comprehensive evaluation. A global score is then estimated for the report as a whole, reflecting both its technical accuracy and domain relevance.

3.2.2. Reports Evaluation Results

Representative examples of expert-evaluated reports are provided in Figures 10 and 11 to illustrate variations in detection accuracy and assessment quality. In the first case (Figure 11), the report received a rating of ‘*Very Good*’, as all landslides within the scene were correctly detected with bounding boxes closely aligned with the ground truth. The

geo-attributes of all three landslides were accurately extracted and presented in the report, leading to high-quality outputs and ‘Very Good’ evaluations for both the analysis and hazard assessment. In contrast, the second case (Figure 10) corresponds to a report rated as ‘Acceptable’. This lower rating was due to the failure to detect the small landslide in the lower-right portion of the image and a slight spatial offset in the bounding box of the upper landslide compared to the ground truth. Table 5 summarizes the distribution of scores across the evaluated report sections and the global evaluation. Results indicate that the majority of the reports (58%) were rated as ‘Very Good’ as the global score, while (30%) were rated as ‘Good’, with (8%) rated as ‘Acceptable’, and (4%) as ‘Poor’. While the strongest section-level performance was observed in Section 3: Analysis and Observation and Section 4: Priority Assessment, where almost two-thirds of the outputs were rated ‘Very Good’ (71% and 70% respectively).

Table 5. Scores for the evaluated sections and global evaluation.

Score	Summary	Analysis and Observation	Priority Assessment	Global Score
Very Good	34%	71%	70%	58%
Good	42%	18%	17%	30%
Acceptable	18%	7%	9%	8%
Poor	6%	3%	4%	4%

1 SUMMARY

Finding: One medium-sized slide is present in the image.

2 GEO-SPATIAL FEATURES

Slide 1

is located in middle-right in an area of **extremely high elevation, moderate slope**, composed of **pelagic limestones, marls, and travertines**, covered by **dense vegetation**.
Conclusion: Dense vegetation and moderate slope likely stabilized the terrain.

3 OBSERVATIONS AND ANALYSIS

- The terrain features a medium-sized slide with steep slopes at high elevation within a clay-marl unit. Dense vegetation across most sectors provides root reinforcement and interception that generally restrain motion and may conceal minor scarps, while the clear separation of the slide from other features indicates localized, independent controls rather than a shared kinematic system or propagating failure.

4 PRIORITY ASSESSMENT

- **Overall risk** is **Medium**, supported by the convergence of steep gradients at high elevation, clay-marl lithology, and the absence of overlap that would indicate kinematic linkage, despite the dense vegetative cover.
- **The probable velocity** is **Moderate**, reflecting translational movements in cohesive materials that can mobilize efficiently on steep faces while remaining partly constrained by matrix cohesion, with localized acceleration where saturation peaks.
- Likelihood is **High** during short, intense rainfall due to rapid pore-pressure rise and softening; **High** throughout prolonged wet periods or snowmelt as sustained moisture degrades shear strength; and **Medium** following vegetation loss, where reduced root reinforcement and interception further lower stability.

Figure 10. Example of a report evaluated as ‘Acceptable’.

1 SUMMARY

Finding: Three small slides are present in the image.

2 GEO-SPATIAL FEATURES

Slide 1

is located in bottom-left in an area of **extremely high elevation, moderate slope**, composed of **sands, sandstones**, and **conglomerates**, covered by **dense vegetation**. Conclusion: Dense vegetation and moderate slope likely stabilized the terrain.

Slide 2

is located in middle-center in an area of **extremely high elevation, moderate slope**, composed of **sands, sandstones**, and **conglomerates**, covered by **dense vegetation**. Conclusion: Dense vegetation and moderate slope likely stabilized the terrain.

Slide 3

is located in middle-left in an area of **extremely high elevation, moderate slope**, composed of **sands, sandstones**, and **conglomerates**, covered by **dense vegetation**. Conclusion: Dense vegetation and moderate slope likely stabilized the terrain.

3 OBSERVATIONS AND ANALYSIS

- The three small slides are located in a terrain with extremely high elevation and moderate slopes. The lithology consists of sands, sandstones, and conglomerates, and all slides exhibit dense vegetation. The absence of proximity and intersections suggests that the slides are not directly adjacent or overlapping, but their spatial arrangement may indicate a shared hydrologic control or slope-parallel seepage.

4 PRIORITY ASSESSMENT

- **Overall risk is Medium**, as the steep relief at extremely high elevation, weakly cemented clastic lithologies, and the absence of proximity and intersections suggest a potential for shallow translational movements.
- **The probable velocity is Moderate**, reflecting the shallow nature of the slides and the moderate slopes, with vegetation limiting acceleration.
- **Likelihood High** during short, intense rainfall due to rapid infiltration and potential pore-pressure rise in permeable clastics; **High** during prolonged wet periods or snowmelt as sustained saturation weakens contacts; and **Medium** following vegetation loss, where reduced root reinforcement and interception would further lower stability.

Figure 11. Example of a report evaluated as ‘Very Good’.

The lower performance observed in Section 1 (Summary) is primarily attributable to errors in the landslide identification stage, as this section directly relies on the outputs of the SegFormer model.

Tables 6 and 7 present the rates of missing information and noisy information as reported by the evaluators. As mentioned above, the missing and noisy content observed in the Summary section primarily originated from identification and localization errors in Stage 1, including false positives (FP) and false negatives (FN), as well as polygon boundary inaccuracies. However, as shown in Table 7, 100% of the reports contained no noisy information in Section 4, and 94% of them were also free of noisy information in Section 3. This interpretation is further supported by the results in Table 8, where, considering only rows with clean inputs, i.e., those for which SegFormer produced an error-free detection (no false positives, no false negatives) fed to the LLM, experts assigned 81% of the reports a “Very Good” global score, 15% a “Good” score, 4% an “Acceptable” score, and 0% a “Poor” score. These findings clearly demonstrate the high quality and reliability of the reports generated by the proposed pipeline. The 4% of reports rated as ‘Poor’ are attributable to Stage 1 detection errors; under error-free detection inputs, the ‘Poor’ rate reduces to 0% (Table 8), confirming that the report generation component does not introduce independent failure modes.

Table 6. Missing info for the evaluated sections.

Missing Info	Summary	Analysis and Observation	Priority Assessment
Yes	21%	19%	16%
No	79%	81%	84%

Table 7. Noisy info for the evaluated sections.

Noisy Info	Summary	Analysis and Observation	Priority Assessment
Yes	11%	3%	0%
No	89%	97%	100%

Table 8. Global score with only clean data (tiles with error-free SegFormer detection).

Score	Percentage
Very Good	81.48%
Good	14.81%
Acceptable	3.70%
Poor	0%

4. Discussion

The results demonstrate that the proposed framework addresses an important challenge in rapid post-disaster assessment by generating structured and technically useful landslide reports. Our enhanced SegFormer exhibits robust performance on a challenging dataset, while the findings also underscore the value of enriching the LLM input with geospatial context, integrating geo-band features such as DEM, slope, NDVI, and lithology to support more domain-relevant interpretations grounded in terrain-context information. Specifically, multi-source data integration operates at two complementary levels within the framework. At the detection level, slope data is fused into the SegFormer encoder via a Squeeze-and-Excitation attention mechanism, enhancing the model performance as confirmed by the ablation results in Table 3, with the slope-augmented model (SDF-SegFormer-B2) improving F1-score from 77.35% to 78.39% and IoU from 63.06% to 63.92% over the RGB-only variant. At the reporting level, all four geo-layers (DEM, slope, NDVI, and lithology) are extracted per slide and provided as metadata to the LLM, grounding its interpretation in physical terrain properties. Nevertheless, the generated reports should be interpreted as qualitative terrain-context summaries intended for rapid post-disaster screening, rather than as quantitative geotechnical hazard assessments or physically based landslide forecasting products.

The framework automates the full workflow from landslide localization to report generation, delivering timely preliminary analyses that may support rapid post-disaster assessment activities. Producing preliminary reports automatically, it mitigates the costs and safety risks of on-site surveys, particularly in areas affected by ongoing instability or adverse weather. In addition, it provides comprehensive geospatial characterization of impacted areas and delivers domain-specific insights, including event interpretation, priority assessment, and actionable recommendations, thereby supporting expert decision-making and intervention planning. The generated outputs can also serve as structured inputs for downstream AI agents tasked with developing mitigation or response plans. The framework showed promising performance when processing complex or low-quality imagery and may potentially be adapted to other natural hazards, highlighting its practical value for rapid screening and disaster-response support.

Overall, the findings reported in Section 3 demonstrate the reliability and generalizability of the proposed pipeline. The performance of the segmentation stage on the unseen

test set, together with the independent expert assessment of the generated reports, suggests that the framework produces outputs that are consistent with the ground truth at both the pixel and the semantic level, which we interpret as a preliminary indication of its reliability within the scope of this study. However, the generated reports remain dependent on the quality of the extracted geo-attributes and detected landslide regions and therefore should be considered as preliminary decision-support outputs rather than substitutes for detailed expert geotechnical assessment or monitoring-based hazard analysis. The framework relies on geospatial layers (DEM, slope, NDVI, and lithology) that are broadly available at a global scale, which in principle makes it transferable to other study areas where comparable inputs can be obtained.

A limitation of this study is that all image tiles in the dataset contain at least one landslide instance, meaning that the model was not exposed to background-only (negative) samples during training or evaluation. Consequently, the reported precision may be optimistic and may not fully reflect performance under real-world conditions, where a large proportion of imagery is landslide-free, potentially leading to higher false positive rates in operational scenarios. This limitation is mainly due to the absence of an independent dataset with similar characteristics to ours, particularly one that includes both landslide annotations and the same multi-source geospatial inputs (DEM, slope, NDVI, and lithology), which are essential for our model. As a result, conducting a more realistic evaluation with negative samples was not feasible within the scope of this study. Additionally, each image tile covers a limited spatial extent ($192\text{ m} \times 192\text{ m}$ at 3 m resolution), which may restrict the ability to capture the broader geomorphological context required for comprehensive landslide analysis. While the integration of multi-source geospatial data enables meaningful local interpretation, certain aspects, such as contributing area, flow paths, and downslope run-out, may extend beyond the boundaries of a single tile. Consequently, scene-level interpretations generated by the model may not fully represent large-scale physical processes and should be interpreted within the spatial limits of the input data.

Future work will focus on enhancing the robustness and generalization of the proposed framework by incorporating additional data sources and extending the spatial reasoning capabilities of the system. In particular, integrating alternative remote sensing modalities such as Synthetic Aperture Radar (SAR) imagery alongside optical data could improve landslide detection performance under challenging conditions, including cloud cover and low visibility. Furthermore, to address the limitations associated with the fixed tile size ($192\text{ m} \times 192\text{ m}$), future developments will explore the integration of spatial context across adjacent tiles by leveraging their geographic coordinates, enabling the LLM to perform more comprehensive reasoning on landslide mechanisms, including contributing areas and downslope processes. Finally, the priority level generated by the framework is intended as a first-order post-event triage indicator to support rapid response efforts, rather than a replacement for continuous monitoring-based hazard assessment. The integration of additional data sources and temporal information is therefore considered a key direction for improving the reliability and operational applicability of the system. In particular, the current framework does not incorporate direct kinematic measurements into the operational prioritization process, as the assessment is derived from geo-attribute extraction and surrounding spatial configuration obtained from post-event data. Future integration of quantitative velocity observations from continuous monitoring systems (e.g., InSAR, GNSS) would strengthen the physical grounding of the framework and enable more robust and precise hazard and priority assessments. Future evaluations could also further quantify expert consistency using independent ratings and inter-rater reliability metrics such as Cohen's or Fleiss' kappa.

5. Conclusions

This study advances rapid landslide intelligence by introducing a hybrid, multi-model framework that links landslide detection, geospatial attribute extraction, and structured, expert-style reporting. The pipeline integrates satellite imagery with terrain and environmental layers to move beyond maps and labels toward concise, decision-oriented summaries suitable for hazard management. By providing context-rich analyses, the framework strengthens evidence-based planning and policy decisions while reducing the cost and safety risks associated with traditional field surveys, particularly in regions characterized by complex terrain or adverse weather conditions. Expert evaluation indicates that the proposed grounding strategy (i.e., enriching the LLM input with geospatial context to anchor the generated narratives to scene-specific, physically meaningful cues) effectively mitigates “hallucinations,” a frequent challenge in vision-language models. The results demonstrate that the LLM produced linguistically and physically consistent reports, with 96% of outputs rated as ‘Good’ or ‘Very Good’ when provided with accurate detection inputs. These findings suggest that the integration of constraint-based prompting and geospatial attributes enables the model to generate expert-style analysis suitable for geohazard contexts. However, the framework’s overall performance is currently sensitive to the initial detection stage. To address this, future work could investigate the use of high-resolution satellite data and the integration of further segmentation and detection architectures. Additionally, additional modalities (e.g., LiDAR, InSAR, rainfall/soil data) can be incorporated. Furthermore, there is significant potential to evolve this framework into an AI-agentic system. By incorporating autonomous workflows, the system could perform more complex, multi-step reasoning, such as cross-referencing historical inventories or integrating real-time rainfall data. Such advancements would further mature the framework into a scalable and reliable tool for early warning and post-disaster recovery in complex terrains.

Author Contributions: M.A.: Writing—original draft, Software, Methodology, Conceptualization, and Data curation. A.R.: Writing—original draft, Software, Methodology, Conceptualization, and Data curation. P.C.: Writing—review and editing, Methodology, Conceptualization, Data curation, and Supervision. M.G.: Writing—review and editing, Software, Methodology, and Visualization. S.B.: Writing—review and editing, Methodology, Conceptualization, Data curation, and Supervision. F.M.: Writing—review and editing, Methodology, Conceptualization, and Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the Space It Up project funded by the Italian Space Agency, ASI, and the Ministry of University and Research, MUR, under contract n. 2024-5-E.0-CUP n. I53D24000060005.

Data Availability Statement: Data will be made available on request.

Use of Artificial Intelligence: Portions of the text were edited with the assistance of ChatGPT 5.1 to improve clarity and stylistic quality.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Xu, Q.; Zhao, B.; Dai, K.; Dong, X.; Li, W.; Zhu, X.; Yang, Y.; Xiao, X.; Wang, X.; Huang, J.; et al. Remote Sensing for Landslide Investigations: A Progress Report from China. *Eng. Geol.* **2023**, *321*, 107156. [[CrossRef](#)]
2. Li, Y.; Fu, B.; Yin, Y.; Hu, X.; Wang, W.; Wang, W.; Li, X.; Long, G. Review on the Artificial Intelligence-Based Methods in Landslide Detection and Susceptibility Assessment: Current Progress and Future Directions. *Intell. Geoengin.* **2024**, *1*, 1–18. [[CrossRef](#)]
3. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4Sense: Reference Benchmark Data and Deep Learning Models for Landslide Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5633017. [[CrossRef](#)]

4. Dai, J.; Dai, X.; Zhang, R.; Ma, J.; Li, W.; Lu, H.; Li, W.; Liang, S.; Dai, T.; Shan, Y.; et al. Using Lightweight Method to Detect Landslide from Satellite Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *135*, 104303. [\[CrossRef\]](#)
5. 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*, 4939. [\[CrossRef\]](#)
6. Wang, B.; Su, J.; Xi, J.; Chen, Y.; Cheng, H.; Li, H.; Chen, C.; Shang, H.; Yang, Y. Landslide Detection with MSTA-YOLO in Remote Sensing Images. *Remote Sens.* **2025**, *17*, 2795. [\[CrossRef\]](#)
7. Soares, L.P.; Dias, H.C.; Garcia, G.P.B.; Grohmann, C.H. Landslide Segmentation with Deep Learning: Evaluating Model Generalization in Rainfall-Induced Landslides in Brazil. *Remote Sens.* **2022**, *14*, 2237. [\[CrossRef\]](#)
8. 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*, 12. [\[CrossRef\]](#)
9. Zhao, W.; Zhang, J.; Cai, J.; Ming, D. Hybrid-SegUformer: A Hybrid Multi-Scale Network with Self-Distillation for Robust Landslide InSAR Deformation Detection. *Remote Sens.* **2025**, *17*, 3514. [\[CrossRef\]](#)
10. Syed, H.M.; Oghaz, M.M.; Saheer, L.B. Semantic Segmentation for Landslide Detection Using Segformer. In *Proceedings of the Artificial Intelligence XLI, Cambridge, UK, 17–19 December 2024*; Bramer, M., Stahl, F., Eds.; Springer Nature: Cham, Switzerland, 2025; pp. 35–45.
11. Xi, L.; Yu, J.; Ge, D.; Pang, Y.; Zhou, P.; Hou, C.; Li, Y.; Chen, Y.; Dong, Y. SAM-CFFNet: SAM-Based Cross-Feature Fusion Network for Intelligent Identification of Landslides. *Remote Sens.* **2024**, *16*, 2334. [\[CrossRef\]](#)
12. Fang, C.; Fan, X.; Wang, X.; Nava, L.; Zhong, H.; Dong, X.; Qi, J.; Catani, F. A Globally Distributed Dataset of Coseismic Landslide Mapping via Multi-Source High-Resolution Remote Sensing Images. *Earth Syst. Sci. Data* **2024**, *16*, 4817–4842. [\[CrossRef\]](#)
13. Areerob, K.; Nguyen, V.-Q.; Li, X.; Inadomi, S.; Shimada, T.; Kanasaki, H.; Wang, Z.; Suganuma, M.; Nagatani, K.; Chun, P.; et al. Multimodal Artificial Intelligence Approaches Using Large Language Models for Expert-Level Landslide Image Analysis. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 2900–2921. [\[CrossRef\]](#)
14. Zhang, Y.; Wu, W.; Qin, Y.; Lin, Z.; Zhang, G.; Chen, R.; Song, Y.; Lang, T.; Zhou, X.; Huangfu, W.; et al. Mapping Landslide Hazard Risk Using Random Forest Algorithm in Guixi, Jiangxi, China. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 695. [\[CrossRef\]](#)
15. Dahal, A.; Lombardo, L. Towards Physics-Informed Neural Networks for Landslide Prediction. *Eng. Geol.* **2025**, *344*, 107852. [\[CrossRef\]](#)
16. Wu, L.; Liu, R.; Ju, N.; Zhang, A.; Gou, J.; He, G.; Lei, Y. Landslide Mapping Based on a Hybrid CNN-Transformer Network and Deep Transfer Learning Using Remote Sensing Images with Topographic and Spectral Features. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103612. [\[CrossRef\]](#)
17. Kuckreja, K.; Danish, M.S.; Naseer, M.; Das, A.; Khan, S.; Khan, F.S. GeoChat: Grounded Large Vision-Language Model for Remote Sensing. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024*; IEEE Computer Society: Los Alamitos, CA, USA, 2024; pp. 27831–27840.
18. He, R.; Zhang, W.; Dou, J.; Jiang, N.; Xiao, H.; Zhou, J. Application of Artificial Intelligence in Three Aspects of Landslide Risk Assessment: A Comprehensive Review. *Rock Mech. Bull.* **2024**, *3*, 100144. [\[CrossRef\]](#)
19. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide Inventory Maps: New Tools for an Old Problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. [\[CrossRef\]](#)
20. Berti, M.; Pizziolo, M.; Scaroni, M.; Generali, M.; Critelli, V.; Mulas, M.; Tondo, M.; Lelli, F.; Fabbiani, C.; Ronchetti, F.; et al. RER2023: The Landslide Inventory Dataset of the May 2023 Emilia-Romagna Meteorological Event. *Earth Syst. Sci. Data* **2025**, *17*, 1055–1074. [\[CrossRef\]](#)
21. Filippini, F.; Iadanza, C.; Vivaldi, V.; Zucca, F.; Meisina, C.; Ferrario, M.F.; Trigila, A. Hybrid Pixel-Based and Object-Based Image Analysis Approach for Landslides Rapid Mapping: The Extreme Rainfall in Emilia-Romagna (Italy) May 2023 Case Study. *Nat. Hazards* **2025**, *121*, 22549–22580. [\[CrossRef\]](#)
22. Ferrario, M.F.; Livio, F. Rapid Mapping of Landslides Induced by Heavy Rainfall in the Emilia-Romagna (Italy) Region in May 2023. *Remote Sens.* **2024**, *16*, 122. [\[CrossRef\]](#)
23. Caleca, F.; Lombardo, L.; Steger, S.; Tanyas, H.; Raspini, F.; Dahal, A.; Nefros, C.; Mărgărint, M.C.; Drouin, V.; Jemec-Auflič, M.; et al. Pan-European Landslide Risk Assessment: From Theory to Practice. *Rev. Geophys.* **2025**, *63*, e2023RG000825. [\[CrossRef\]](#)
24. Sani, F.; Bonini, M.; Piccardi, L.; Vannucci, G.; Donne, D.D.; Benvenuti, M.; Moratti, G.; Corti, G.; Montanari, D.; Sedda, L.; et al. Late Pliocene–Quaternary Evolution of Outermost Hinterland Basins of the Northern Apennines (Italy), and Their Relevance to Active Tectonics. *Tectonophysics* **2009**, *476*, 336–356. [\[CrossRef\]](#)
25. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Proceedings of the Advances in Neural Information Processing Systems, San Diego, CA, USA, 6–14 December 2021*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 12077–12090.
26. Jocher, G.; Qiu, J. Ultralytics YOLO11, version 11.0.0. 2024. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 28 May 2026).

27. Lv, W.; Zhao, Y.; Chang, Q.; Huang, K.; Wang, G.; Liu, Y. RT-DETRv2: Improved Baseline with Bag-of-Freebies for Real-Time Detection Transformer 2024. *arXiv* **2024**, arXiv:2407.17140.
28. Mashari, S.; Solaimani, K.; Omidvar, E. Landslide Susceptibility Mapping Using Multiple Regression and GIS Tools in Tajan Basin, North of Iran. *Environ. Nat. Resour. Res.* **2012**, *2*, 43–51. [[CrossRef](#)]
29. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA: Washington, DC, USA, 1973.
30. Özyavuz, M.; Bilgili, C.; Salıcı, A. Determination of Vegetation Changes with NDVI Method. *J. Environ. Prot. Ecol.* **2015**, *16*, 264–273.
31. Nadhavajhala, S.; Tong, Y. *Rubra-Mistral-7B-Instruct-v0.3*; HuggingFace: New York, NY, USA, 2024. [[CrossRef](#)]
32. Cruden, D.M.; Varnes, D.J. *Landslide Types and Processes*, Transportation Research Board; U.S. National Academy of Sciences, Special Report; Transportation Research Board, National Research Council: Washington, DC, USA, 1996; Volume 247, pp. 36–75.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.