Rapid Landslide Mapping During the 2023 Emilia-Romagna Disaster: Assessing Automated Approaches with Limited Training Data ~~for Emergency Response~~

~~*Dal Seno N.[1], Ciccarese G.[1], Evangelista D.[2], Piccolomini E.[2], Berti M.[1]*~~

Nicola Dal Seno[1], Giuseppe Ciccarese [1], Davide Evangelista[2], Elena Loli Piccolomini[2], Alessandro Corsini[3], Matteo Berti[1]*

[1] Department of Biological, Geological, and Environmental Sciences (BiGeA), University of Bologna, via Zamboni 67, Bologna, Italy

[2] Department of Informatics: Science and Engineering (DISI), University of Bologna, Bologna, Italy

[3] Department of Chemical and Geological Sciences, University of Modena, Modena, Italy

*Corresponding author

email: nicola.dalseno@unibo.it

full postal address: Via Zamboni, 67 - 40126 Bologna

~~Abstract~~

The catastrophic rainfall events of May 2023 in the Emilia-Romagna region, Italy, triggered ~~80997~~more than 80,000 landslides~~ (https://doi.org/~~, as documented in a publicly available inventory (10.5281/zenodo.13742643~~, Pizziolo et al., 2024), placing~~), and placed extraordinary demands on emergency response systems. One of the most critical emergency tasks was landslide mapping, which was carried out manually and required substantial time and effort. This study ~~explores~~investigates the potential of automated landslide mapping to support rapid disaster response~~,~~ by evaluating two deep learning models, U-Net and SegFormer, under realistic emergency constraints, including ~~minimal~~limited training data.

The models were trained ~~exclusively in~~ using data from one affected ~~area (~~municipality, Casola Valsenio~~)~~ (84 km²), and tested on three additional municipalities~~ (~~, Predappio~~,~~ (91 km²), Modigliana~~,~~ (101 km²), and Brisighella~~) with varying~~ (194 km²), characterized by different geological settings. To ~~reflect~~represent a range of operational scenarios, we tested seven combinations of input ~~layers, progressively increasing in complexity~~data, from post-event Sentinel-2 imagery alone to ~~full~~the integration of high-resolution aerial imagery, ~~NDVI~~vegetation change maps, and slope data.

~~Results show that both~~ Both models achieved comparable segmentation performance, with SegFormer ~~displaying~~showing greater robustness to variations in input ~~layers~~data and geological conditions, while U-Net was more sensitive but occasionally more accurate ~~with rich~~when richer inputs~~.~~ were available. Both models successfully identified landslides, but ~~encountered difficulties~~showed limitations in shadowed ~~zones~~areas, cultivated fields, and geologically distinct terrains. A ~~key~~major limitation emerged in the Brisighella ~~FAA, a sector dominated by Blue Clay formations~~area, where poor generalization was ~~traced to the lack of~~associated with the dominance of Blue Clay formations and the limited lithological diversity ~~in~~of the training data~~, highlighting~~. These results highlight the ~~need for~~importance of geologically balanced ~~training~~ datasets~~.~~ for improving model transferability.

Definizione stile: Titolo 2

~~Despite these challenges~~Overall, the study confirms the operational value of automated landslide mapping as a first-pass tool~~. Both models delivered accurate baseline maps that could support manual validation and prioritization in time-critical scenarios. The findings suggest that AI-based mapping could become a key component of emergency management protocols. While~~ for emergency response. Although manual revision remains ~~essential, these tools offer~~necessary, both models produced reliable baseline maps that can support validation and help prioritize interventions in time-critical situations, offering a scalable~~,~~ and time-efficient ~~solution~~approach to the ~~increasing demand~~growing need for rapid and spatially detailed hazard information.

1. Introduction

Rapid and accurate landslide mapping over large areas is a challenging task, particularly in the context of regional-scale rainfall or earthquake events (Iverson et al. 2015; Casagli et al. 2016; Robinson et al., 2017; Holbling et al. 2017; Mondini et al., 2021). This task is crucial for implementing timely and effective response measures, assessing the extent of impact, and planning for recovery and mitigation efforts. The complexities of rapid landslide mapping vary depending on several factors including the extent of the affected area, the availability of cloud-free imagery, the precision needed for the mappings, and the difficulties of detecting landslides due to shadows, vegetation cover, or weak geomorphic evidences (Mondini et al. 2019; Amatya et al., 2023). Typically, these complexities are addressed by expert geologists through the visual interpretation of satellite or aerial imagery (Guzzetti et al., 2012; Scaioni et al., 2014; Ferrario and Livio 2023). These experts are skilled at identifying subtle variations in the terrain and can integrate complex contextual knowledge, including field reports, personal accounts, and specific geological information, to create highly accurate landslide maps. However, manual mapping is both time-consuming and inherently subjective, which are significant limitations during ~~a crisis~~an emergency (Novellino et al. 2024).

Before the widespread adoption of deep learning, automated landslide mapping approaches evolved through several methodological stages. Early methods relied mainly on pixel-based techniques, such as thresholding of spectral indices or terrain parameters, which were simple to implement but highly sensitive to noise and illumination conditions. These approaches were later complemented by object-based image analysis (OBIA), which segments images into meaningful objects and classifies them based on spectral, spatial, and contextual features (Blaschke 2010; Martha et al. 2010). OBIA-based methods represented a substantial improvement by incorporating shape, texture, and neighborhood information, and were widely applied to landslide detection from high-resolution imagery. However, both pixel- and object-based approaches typically depend on manually designed rules or handcrafted features, limiting their transferability across different events, sensors, and geological settings.

More recently, machine-learning and deep-learning approaches have progressively replaced rule-based systems by enabling data-driven feature extraction and end-to-end learning. Automated landslide recognition techniques using convolutional neural networks (~~CNN) algorithms are emerging~~CNNs) have emerged as promising alternatives to manual ~~methods~~mapping (Sameen and Pradhan 2019~~,~~; Chen et al. 2018~~,~~; Ye et al. 2019~~,~~; Tang et al. 2022~~,~~; Ji 2020). ~~Recent~~Numerous studies have demonstrated their effectiveness in real-world ~~cases.~~post-disaster scenarios. For ~~instance~~example, Meena et al. (2021) applied a deep learning approach for rapid landslide mapping in India following extreme monsoon rainfall, ~~showcasing the potential of these techniques in post-~~

~~disaster conditions.~~ while Prakash et al. (2021) introduced a ~~new strategy using a~~ generalized ~~convolutional neural network, aiming to develop a more versatile method applicable across various geographical~~CNN framework designed to improve applicability across different geographic contexts. ~~Building on this,~~ Prakash and Manconi (2021) further demonstrated the ~~practical utility~~operational value of ~~these methods~~CNN-based mapping by rapidly ~~mapping~~delineating landslides triggered by severe storm events. Beyond individual case studies, recent research has focused on transferability, reproducibility, and systematic comparison of methods. Prakash et al. (2020) explicitly compared deep-learning models with traditional machine-learning approaches for landslide mapping from Earth observation data, highlighting the advantages of deep architectures in complex environments. In parallel, benchmark datasets such as Landslide4Sense (Ghorbanzadeh et al. 2022), HR-GLDD (Meena et al. 2023), and the CAS Landslide Dataset (Xu et al. 2024) have been released to promote standardized evaluation across multiple events, sensors, and geographic regions. These efforts have significantly advanced the field by enabling more robust inter-comparisons and by revealing persistent challenges related to class imbalance, spatial heterogeneity, and generalization across geological domains.

In parallel with CNN developments, transformer-based architectures have recently been introduced in landslide mapping to better capture long-range spatial dependencies and multi-scale contextual information. Unlike CNNs, which primarily exploit local convolutional kernels, transformers rely on self-attention mechanisms that allow each pixel (or patch) to attend to a broader spatial context. This property is particularly relevant for landslide detection, where the geomorphological setting, slope connectivity, and spatial coherence of failures extend beyond local neighborhoods. Hybrid CNN–Transformer models and pure transformer-based architectures have shown promising results in improving robustness and generalization across heterogeneous landscapes (Chen et al. 2023; Wu et al. 2024). These characteristics are especially attractive for rapid response applications, where models must be deployed with limited training data and applied to large, spatially diverse areas under operational constraints.

Despite these advances, a ~~severe storm event.~~ substantial gap remains between methodological developments and their practical deployment during emergency response. Most deep-learning studies rely on relatively large and well-balanced training datasets, often adopting train-to-validation ratios such as 80:20 or 70:30 (Meena et al. 2023). In contrast, during an ongoing crisis only a small fraction of the affected area is typically mapped and annotated in the first hours to days, making extensive training datasets unrealistic. This limitation is often exacerbated by delays in acquiring cloud-free post-event imagery, inconsistencies in pre-event reference data, and the operational need to deploy models rapidly with minimal manual intervention. Furthermore, early training samples are frequently not representative of the full variability of landslide processes and environmental conditions, including different landslide types, lithologies, land-cover patterns, and illumination or shadow effects. As a result, models trained on limited and potentially biased early annotations may generalize poorly when applied to the broader affected region.

~~These studies emphasize the growing importance of CNNs in automated landslide mapping and their potential for practical implementation in disaster management. However, despite their proven effectiveness, a notable gap remains between theoretical applications and their real-world deployment during emergencies. Typically, these studies employ a large train-to-validation ratio (usually 80:20 or 70:30; Meena et al 2023), which does not mirror the limited data availability for training models~~

ha formattato: Tipo di carattere: Times New Roman, 12 pt, Legature: Nessuna

~~in real-time crisis situations. Furthermore, the training process frequently does not account for the variety of landslide types and the spectrum of geological conditions present in actual situations. This oversight is critical, as the complexity of natural environments play a crucial role in the effective application of these technologies in emergency contexts.~~

In this study, we utilize the landslide inventory from the May 2023 crisis in the Emilia-Romagna region of Italy (Berti et al. 2025) to assess the effectiveness and limitations of automated mapping algorithms in a practical setting. We specifically focus on the performance of the U-Net and SegFormer algorithms in identifying landslides triggered by the event, despite the challenges of very limited training data. The study also considers the variety of landslide types and geological conditions prevalent in the area, which impacted the disaster response. The main goal of this research is to determine whether advanced deep-learning methods can effectively replace manual mapping in managing large-scale landslide disasters, thus bridging the gap between academic research and real-world application in emergency management.

2. The Romagna May 2023 disaster

2.1 Overview of the disaster

In May 2023, the Emilia-Romagna region of Italy was hit by unprecedented meteorological events, as detailed by Berti et al. (2025) and Foraci et al~~.,~~. (2023). Two major rainfall episodes, from May 1-3 and May 16-17, led to severe floods and landslides primarily affecting the eastern part of the region (Fig. 1a). The first event (May 1-3) was driven by a stationary low-pressure system which delivered over 200 mm in 48 hours. The second event (May 16-17), featured a similar low-pressure system enhanced by bora winds and moist air from the Mediterranean, resulting in rainfall exceeding 250 mm in the same duration. The cumulative impact of these episodes resulted in more than 500 mm of rainfall over two weeks, nearly half the region's annual climatic precipitation (Fig. 1a). The return period for these combined rainfall events was estimated to exceed 1000 years (Brath et al~~.,~~. 2023), highlighting the unprecedented nature of this occurrence. The two cyclonic events led to extensive flooding, submerging 540 km² of land, and triggered more than 80,000 landslides across approximately 1000 km^2 (Fig. 1b). This widespread devastation impacted thousands of buildings, roads, and other infrastructure, with the estimated damages amounting to approximately 9 billion euros.

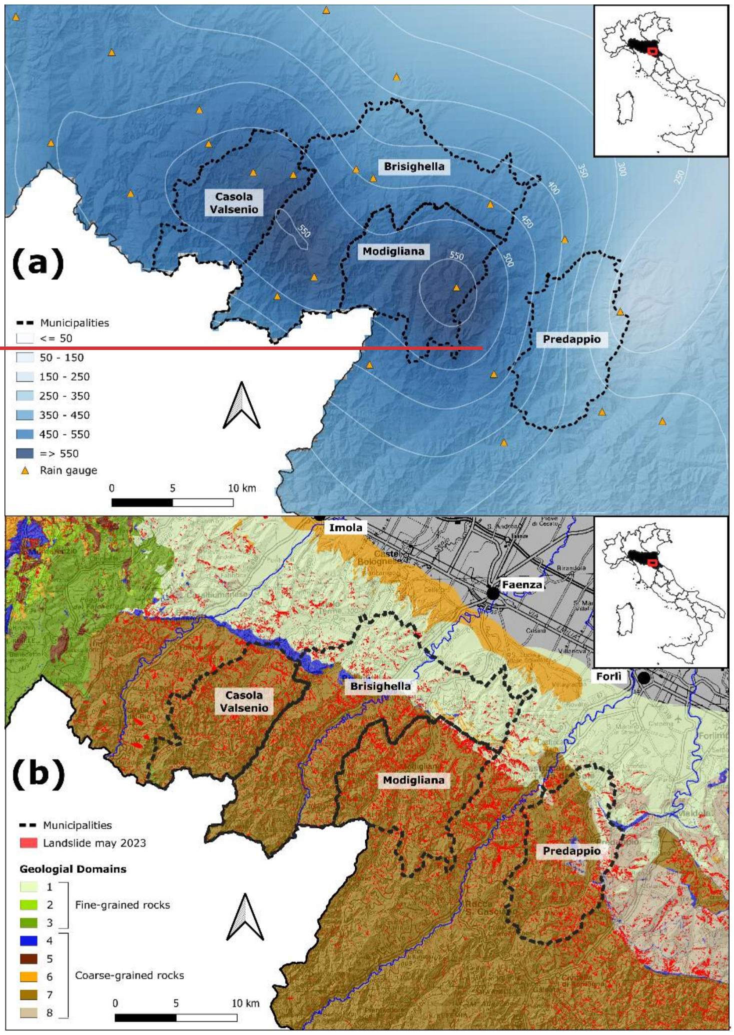
Immediately after the disaster, we initiated the development of a detailed inventory map of the landslides caused by the event. This map was manually created using high-resolution aerial imagery taken only one week following the second rainfall event. The aerial images, which include four bands (RGB+NIR) and have a resolution of 0.2 m, allowed for the accurate identification and classification of the landslides. The data derived from this mapping effort are described in Berti et al. (2025) and are available for public download in shapefile format (DOI: 10.5281/zenodo.13742643, Pizziolo et al., 2024).

2.2 Motivation of the work

The visual identification of the landslides from May 2023 proved to be a challenging and time-intensive process. It took six months for twelve expert geologists to identify and delineate all the landslides within the 1000 km² area affected by the disaster. Despite these challenges, this approach was deemed the best available method at the time to meet the Civil Protection agency's specific requirements for accuracy, precision, and oversight during the emergency response coordination.

As this manual mapping progressed, we explored various automated mapping techniques to speed up the process and reduce the workload. Initial trials using NDVI methods and U-Net in a specific sample area produced promising results, demonstrating that deep-learning models can effectively replicate manual mapping efforts (Berti et al., ~~2024a~~2026). However, these preliminary analyses were conducted in a test area (the Casola Valsenio municipality, Fig. 1), which covers only about 10% of the total area impacted by the event. Consequently, several critical questions remain unanswered: i) Can automated models trained on a small area be effectively used for rapid mapping of larger regions? ii) How do models trained in specific geological conditions perform across different geological settings? iii) Could more sophisticated deep-learning methods surpass U-Net in improving outcomes?

Addressing these questions is crucial to determine whether emerging automated methods can enhance the efficiency of landslide mapping in future crisis scenarios, thereby potentially reducing critical response times. In this study, we simulate landslide mapping during a crisis, taking into account the limited data typically available and the accessible information under such conditions. This simulation is grounded in our firsthand experience of the real emergency during the May 2023 event, which provided us with valuable insights into the practical challenges and data limitations encountered during such crises.


(a)
Brisighella
Casola Valsenio
Modigliana
Predappio
550
500
450
400
350
300
250
Municipalities
<= 50
50 - 150
150 - 250
250 - 350
350 - 450
450 - 550
=> 550
Rain gauge
0
5
10 km
(b)
Imola
Faenza
Forlì
Brisighella
Casola Valsenio
Modigliana
Predappio
Municipalities
Landslide may 2023
Geologial Domains
1
2
3
Fine-grained rocks
4
5
6
7
8
Coarse-grained rocks
0
5
10 km

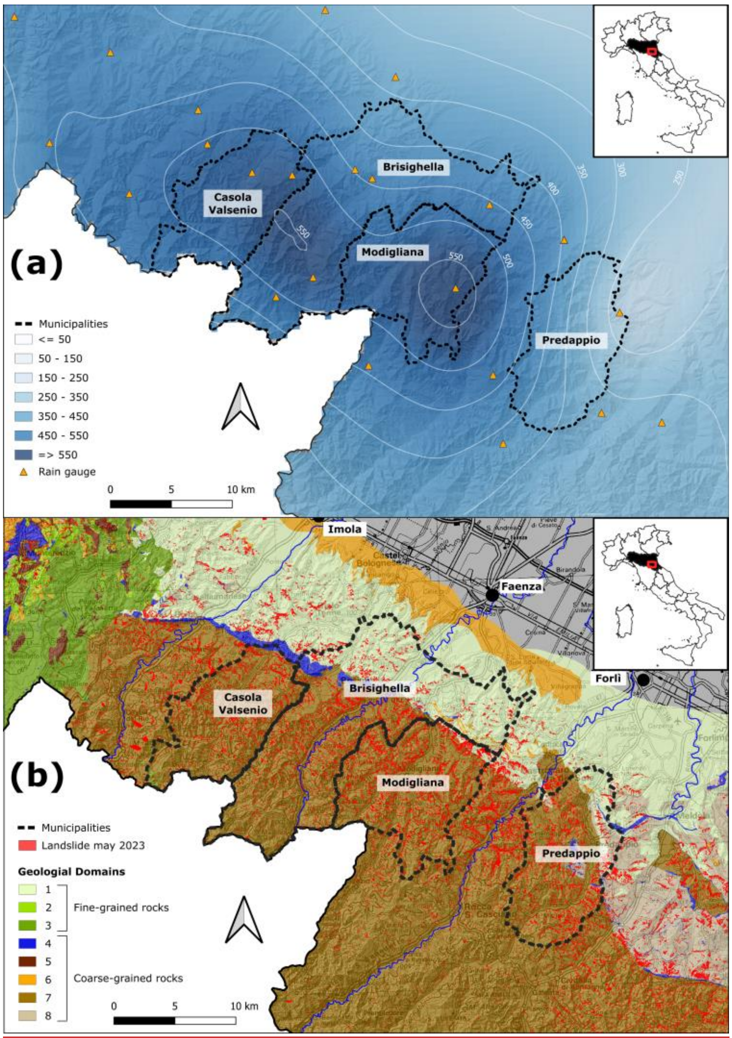


Figure 1. Overview of the Effects of the Extreme Rainfall Event in Romagna, May 2023.
a) Cumulative precipitation ~~map~~ from ~~May~~ 1 to ~~May~~ 17~~,~~ May 2023~~.~~ derived from the regional rain-gauge network operated by Regione Emilia-Romagna. **b**) Manually mapped landslide distribution across the study areas, categorized by ~~geological formations as per~~lithological units (Units 1-8) following the classification proposed by Berti et al. (2025; Table 2).

ha formattato: Colore carattere: Testo 1, Legature: Standard + Contestuali

Formattato: SpazioPrima: 0 pt

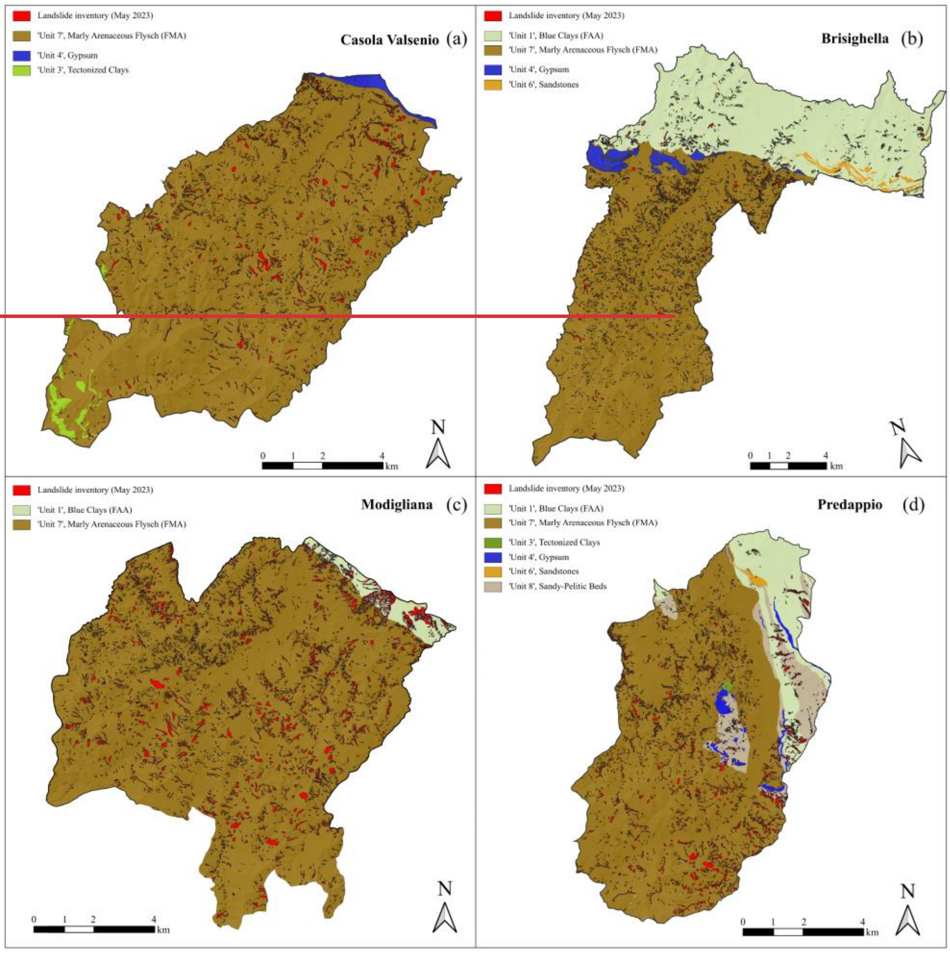
3. Study area

~~Automated methods for landslide mapping were applied in the four municipalities shown in Figure 2. These areas, situated in the eastern part of the Emilia-Romagna region (Fig. 1), were significantly affected by the meteorological event of May 2023. The Casola Valsenio area was chosen for training models, whereas Brisighella, Modigliana, and Predappio served as validation sites. In their study, Berti et al. (2024a) used Casola Valsenio to train a conventional U-Net model, and compared the results with those obtained using the traditional NDVI change detection method. We opted to continue using this area for training to ensure consistency with previous research. The other three areas were selected because together they encompass approximately 35% of the regions most impacted by landslides and represent about 25% of the recorded 80,997 landslide events. Moreover, these municipalities exhibit varied geological conditions, with both similarities and differences to Casola Valsenio, thereby offering a comprehensive test of the models' adaptability to diverse terrains and landforms.~~

~~To represent geological variability in a way that is both meaningful and operational for the analysis, we adopted the lithological classification introduced by Berti et al. (2025), which groups the over 600 geological formations of the Emilia-Romagna region into eight lithological units (Fig. 1b). This classification combines lithological characteristics with structural domains, acknowledging that the same rock type may exhibit different mechanical properties depending on its tectonic setting. Units 1 to 3 consist of fine-grained materials (e.g., clays, marls, tectonized shales), while units 4 to 8 comprise coarse-grained rocks (e.g., sandstones, flysch, conglomerates). This simplified scheme captures key differences in the weathering behavior and mechanical response of the rock masses, which are known to influence the type and frequency of landslides. Fine-grained units generally produce cohesive soils prone to earth slides and flows, whereas coarse-grained units generate granular soils more susceptible to debris slides and debris flows. This categorization aligns with the "earth" and "debris" material types defined in the Cruden and Varnes (1996) landslide classification and is used throughout this study to interpret landslide patterns in relation to geological setting.~~

~~**Table 1.** Simplified classification of the geological formations in the Emilia-Romagna region into eight litho-structural units, adapted from Berti et al. (2025). The classification is based on lithological composition and structural domain. Units 1 to 3 predominantly include fine-grained rocks, while units 4 to 8 are composed mainly of coarse-grained lithologies.~~

Unit ID	Lithology	Domain	Structural position	Geological Age
1	Clays, silty clays, and marly clays	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene
2	Marls and marly clays	Epiligurian	Wedge-top basins	Oligocene to Miocene
3	Clay shales, clay breccias, tectonized clays, olistostromes	Ligurian	Accretionary wedge	Cretaceous to Eocene
4	Massive rocks: basalts, serpentines, limestones, arenites	Ligurian, Epiligurian	Accretionary wedge Wedge-top basins	Cretaceous to Miocene
5	Flysch rocks made of rhythmic alternations of sandstones, limestones, pelites, and shales	Ligurian, Epiligurian	Accretionary wedge Wedge-top basins	Cretaceous to Eocene
6	Weakly cemented sandstones and conglomerates	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene
7	Flysch rocks made of rhythmic alternations of sandstones and pelites	Tuscan-Umbrian	Inner Foredeep	Miocene
8	Weakly cemented sandstones with interbedded pelitic layers	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene



~~**Figure 2.** Schematic geological map of the municipalities of Casola Valsenio (**a**), Brisighella (**b**), Modigliana (**c**) and Predappio (**d**). FMA = Marnoso-Arenacea Formation, FAA = Blue Clays Formation.~~

Casola Valsenio

The municipality of Casola Valsenio is situated 30 km southwest of Imola (Fig. 1), covering an area of 84 km². It is predominantly known for its agricultural activities, including the cultivation of herbs, fruit orchards, and beekeeping. The underlying geology of the region is predominantly made up of the 'Unit 7' ~~shows in Fig. 2a~~ mainly composed ~~by~~of the Marnoso-Arenacea Formation (FMA) (Unit 7 in the lithological classification of Berti et al., 2025; Table 2), a Tertiary flysch characterized by a finely interbedded sequence of marl and sandstone layers. This sequence was deposited in a deep marine foredeep basin during the Miocene epoch (Amy et al., 2006). The formation is characterized by a variable Arenites/Pelites (A/P) ratio, which reflects variations in depositional environments within the basin. In Casola Valsenio, nearly the entire area features FMA with an A/P ratio ranging from 1/3 to 3/1. The northern sector is distinguished by predominantly sandstone-~~ricce~~rich layers (A/P > 3) along with minor outcrops of Tectonized Clays (FPP) and Messinian Gypsum (FGY) formations. Before May 2023, the Geological Survey had documented 730 ancient, inactive landslides, primarily rock-block slides on cataclinal slopes that align with bedding planes. Following the May 2023 disaster, this figure dramatically increased to 5572 new landslides (Fig. 2a). ~~Of these,~~ Landslide types and counts reported here are derived from the official RER2023 inventory (Berti et al., 2025), which is openly available as a vector dataset on Zenodo (Pizziolo et al., 2024; https://doi.org/10.5281/zenodo.13742643). The post-event landslide inventory in Casola Valsenio is dominated by debris slides ~~(DS)~~ and debris flows ~~(DF) accounted for the vast majority, comprising 83% (4,543 events) and 14% (789 events) of the total, respectively. Typically, these were initiated~~, mainly triggered by shallow failures of the thin soil ~~layer (averaging 0.5-1 meters in depth) that overlies~~cover overlying the Marnoso-Arenacea bedrock (Fig. 3b~~-c~~-d~~). In line with Berti et al. (2025), DF events were further classified into~~; Table 1). Both long-~~runout (DF1, 334 events)~~ and limited-runout ~~(DF2, 445 events), while DS events included 63 occurrences of limited-runout type (DS2, 1%). Additionally, 17 earth~~debris flows are observed, reflecting differences in slope morphology and vegetation. Earth movements ~~were recorded, including earth flows (EF, 11 events; 0.2%) and earth slides (ES, 6 events; 0.1%). The area also witnessed several large~~are rare, whereas rock-block slides ~~on cataclinal slopes~~, although less frequent, are locally significant due to their larger size and damage potential (Fig. 3a). ~~Although these accounted for only 2% of the total landslides (110 events), their size was significantly greater than that of the debris slides and flows, resulting in extensive damage.~~ The May 2023 disaster notably disrupted local infrastructure, blocking the main road and cutting off the southern region for weeks. Nearly 70 homes and 200 roadways were severely affected.

Brisighella

The municipality of Brisighella borders Casola Valsenio to the east (Fig. 1) and covers 194 km². The area is primarily composed of the 'Unit 7' composed by Marnoso-Arenacea Formation (FMA) (Fig. 2b). The northern part of the municipality, approximately 75 km² in size, consists mainly of the Blue Clays Formation (FAA) (Fine-grain rocks 'Unit 1' in Fig. 1b), which is bordered by a narrow strip of gypsum (Fig. 2b). Before the event, the Geological Survey had documented 2,100 landslides in the area, mostly ancient and inactive rock-block and complex slides. Following the disaster, an additional 6342 landslides were recorded. Landslide processes vary markedly with lithology in this area. Debris slides prevail in areas underlain by the Marnoso-Arenacea Formation, while earth slides and earth flows are more common in the Blue Clays sector, where badland morphologies are widespread (Table 1). Debris flows occur with variable runout, and rock-block slides are relatively rare but locally impactful.

~~Of these, 61% were debris slides (DS) (3,890 events), while debris flows (DF) accounted for 16% of the total (999 events), 6% of the with a long-runout (DF1) and 10% with limited-runout (DF2). Earth slides (ES) and earth flows (EF), typically associated with badlands in clayey terrains, represented 22% of the total (1,428 events). These types of movements, typical of badlands, are characterized by the progressive erosion and mobilization of clay slopes. Lastly, rock-block slides accounted for 0.4% of the recorded events (25 occurrences). Despite their limited number, these movements are~~

~~significant due to their larger size and potential danger. In the municipal territory, landslides blocked key access routes and damaged 66 homes and 261 roadways.~~

Modigliana

The municipality of Modigliana is located approximately 15 km to the east of Casola Valsenio (Fig. 1), and is one of the area most seriously affected by the event. The area spans 101 km² and is primarily composed of the Marnoso-Arenacea Formation, FMA ('Unit 7') with an A/P ratio ranging from 1/3 to 3/1 (Fig. 2c~~), with a~~). A small portion of Blue Clays, FAA (Unit 1) ~~outcropping~~outcrops in the northeast. Before the disaster, 795 landslides were documented, mostly complex and rock-block slides. After the May 2023 event, the number of landslides jumped to 6974, drastically reshaping the landscape. ~~Of these, debris slides were the most common, accounting for 77% of the total (5357 events). Debris flows (DF) represented 19%, divided between long-runout (DF1, 573 events) and limited-runout (DF2, 733 events) types, while rock-block slides accounted for 1% of the total (69 events). Earth slides (ES) and earth flows (EF) made up 3.5% (242 events). The disaster blocked the main access road for weeks, affecting 51 homes and 155 roadways, and left the community to face long-term recovery challenges~~The landslide inventory in Modigliana is largely dominated by debris slides, with debris flows representing a secondary but relevant process (Table 1). Both long- and limited-runout debris flows contributed to slope and channel instability, whereas rock-block slides and earth movements occur only sporadically. This pattern reflects widespread shallow instability within the Marnoso-Arenacea Formation.

Predappio

The municipality of Predappio is situated around 30 km east of Casola Valsenio (Fig. 1) and spans an area of 91 km². The geological framework of Predappio is mainly characterized by the Marnoso-Arenacea Formation, FMA ('Unit 7') with an A/P ratio ranging from 1/3 to 3/1. The central part of the municipality also features the Colombacci Formation ('Unit 8' in Fig. 1b), interspersed with gypsum from the Gessoso-Solfifera Formation ('Unit 3'). Towards the northeast, the landscape transitions to the Blue Clays formation ('Unit 1'), interbedded with deposits of the Colombacci Formation and several occurrences of Pliocene sandstones (Borello Sandstones, 'Unit 7').

~~Before May 2023, 840 landslides had been recorded, primarily involving rock-block slides and debris flows. Following the disaster, this number escalated to 6832, significantly altering the landscape. Of these, 80% were debris slides (5.476 events), while debris flows (DF) accounted for 11% (750) of the total including 419 long-runout events (DF1) and 331 limited-runout events (DF2). Rock-block slides (RS), although only 0.5% of the recorded events (31 occurrences), were notable for their significant size and potential impact. Earth slides (ES) and earth flows (EF) represented 8% of the total (551 events).~~

Debris slides, here, represent the most common landslide type, accompanied by numerous debris flows with variable runout behavior (Table 1). Rock-block slides are infrequent but potentially severe, while earth slides and earth flows are more common than in purely sandstone-dominated areas, reflecting the local presence of clay-rich lithologies. The area experienced weeks of isolation due to several landslides blocking the main road, impacting 19 homes and damaging 138 roadways, again with long-lasting consequences.

ha formattato: Tipo di carattere: Times New Roman

Formattato: SpazioPrima: 0 pt, Dopo: 8 pt, Interlinea: multipla 1,08 ri

Table 1. Summary of the landslides mapped in the four study municipalities after the May 2023 Emilia-Romagna event, including total counts and breakdown by landslide type. DS = debris slides; DF = debris flows; DF1 = long-runout debris flows; DF2 = limited-runout debris flows; ES = earth slides; EF = earth flows; RS = rock-block slides. Data are from the RER2023 landslide inventory (Berti et al., 2025) and its open-access release (Pizziolo et al., 2024).

Municipality	Tot. Landslides	Debris Slides (DS)	Debris Flows (DF)	Earth Slides (ES)	Earth Flows (EF)	Rock Slides (RS)
Casola Valsenio	5572	4543	789	6	11	110
Brisighella	6342	3890	999	835	593	25
Modigliana	6974	5357	1306	152	90	69
Predappio	6832	5476	750	426	117	31

Automated methods for landslide mapping were applied in the four municipalities shown in Figure 2 (Casola Valsenio, Brisighella, Modigliana and Predappio). These areas, situated in the eastern part of the Emilia-Romagna region (Fig. 1), were significantly affected by the meteorological event of May 2023. The Casola Valsenio area was chosen for training models, whereas Brisighella, Modigliana, and Predappio served as validation sites. In their study, Berti et al. (2026) used Casola Valsenio to train a conventional U-Net model, and compared the results with those obtained using the traditional NDVI change detection method. We opted to continue using this area for training to ensure consistency with previous research. The other three areas were selected because together they encompass approximately 35% of the regions most impacted by landslides and represent around 25% of the recorded 80,997 landslide events (20148). Moreover, these municipalities exhibit varied geological conditions, with both similarities and differences to Casola Valsenio, thereby offering a comprehensive test of the models' adaptability to diverse terrains and landforms.

To represent geological variability in a way that is both meaningful and operational for the analysis, we adopted the lithological classification introduced by Berti et al. (2025), which groups the over 600 geological formations of the Emilia-Romagna region into eight lithological units (Table 2, Fig. 1b). This classification combines lithological characteristics with structural domains, acknowledging that the same rock type may exhibit different mechanical properties depending on its tectonic setting. Units 1 to 3 consist of fine-grained materials (e.g., clays, marls, tectonized shales), while units 4 to 8 mainly comprise coarse-grained rocks (e.g., sandstones, flysch, conglomerates). This simplified scheme captures key differences in the weathering behavior and mechanical response of the rock masses, which are known to influence the type and frequency of landslides. Fine-grained units generally produce cohesive soils prone to earth slides and flows, whereas coarse-grained units generate granular soils more susceptible to debris slides and debris flows. This categorization aligns with the "earth" and "debris" material types defined in the Cruden and Varnes (1996) landslide classification and is used throughout this study to interpret landslide patterns in relation to geological setting.

Table 2. Simplified classification of the geological formations in the Emilia-Romagna region into eight litho-structural units, adapted from Berti et al. (2025). The classification is based on lithological composition and structural domain. Units 1 to 3 predominantly include fine-grained rocks, while units 4 to 8 are composed mainly of coarse-grained lithologies.

Unit ID	Lithology	Domain	Structural position	Geological Age
1	Clays, silty clays, and marly clays	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene
2	Marls and marly clays	Epiligurian	Wedge-top basins	Oligocene to Miocene
3	Clay shales, clay breccias, tectonized clays, olistostromes	Ligurian	Accretionary wedge	Cretaceous to Eocene
4	Massive rocks: basalts, serpentines, limestones, arenites	Ligurian, Epiligurian	Accretionary wedge Wedge-top basins	Cretaceous to Miocene
5	Flysch rocks made of rhythmic alternations of sandstones, limestones, pelites, and shales	Ligurian, Epiligurian	Accretionary wedge Wedge-top basins	Cretaceous to Eocene
6	Weakly cemented sandstones and conglomerates	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene
7	Flysch rocks made of rhythmic alternations of sandstones and pelites	Tuscan-Umbrian	Inner Foredeep	Miocene
8	Weakly cemented sandstones with interbedded pelitic layers	Padano-Adriatic	Outer Foredeep	Pliocene to Pleistocene

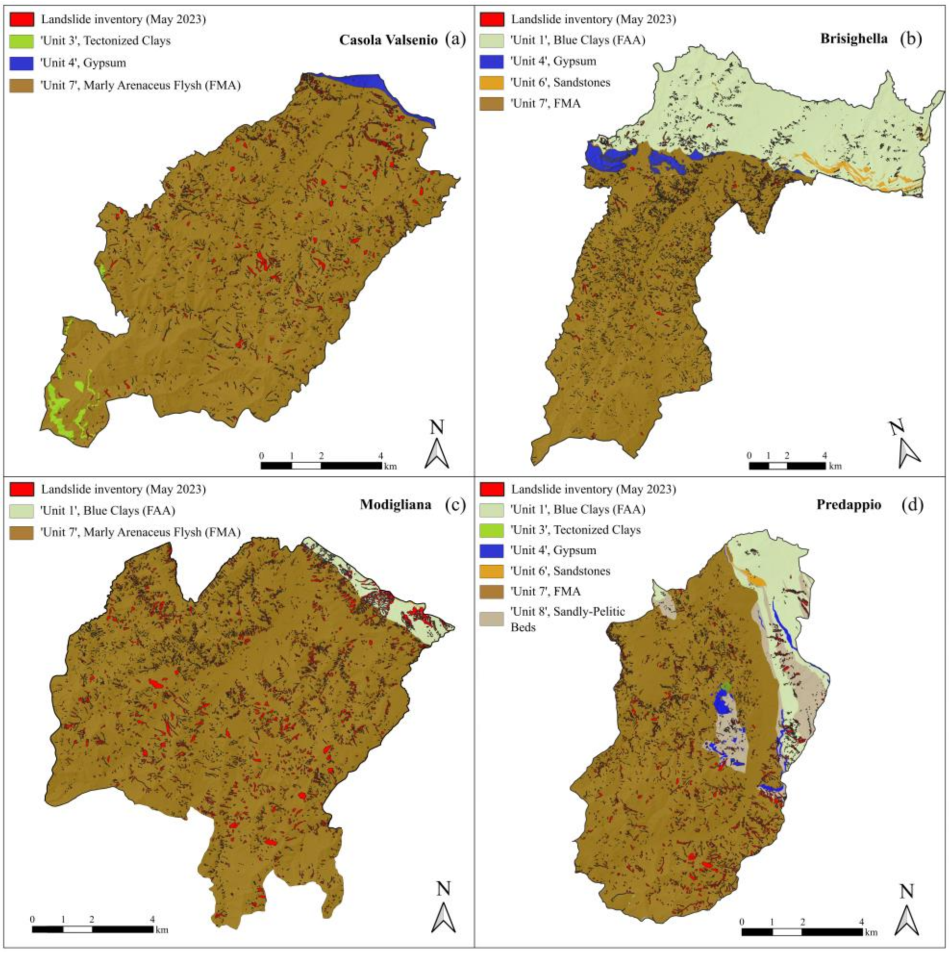
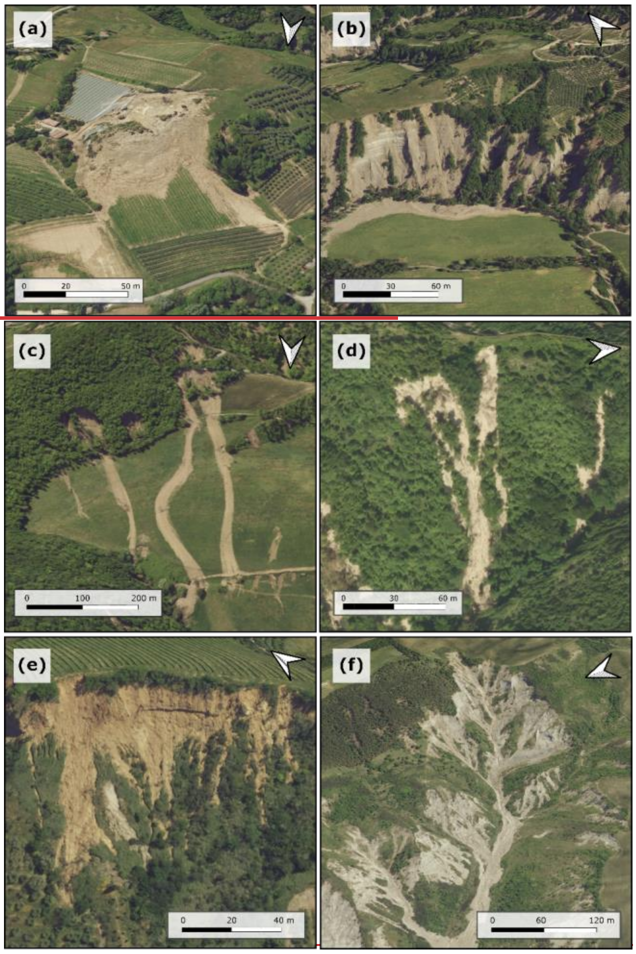


Figure 2. Schematic geological map of the municipalities of Casola Valsenio (**a**), Brisighella (**b**), Modigliana (**c**) and Predappio (**d**). FMA = Marnoso-Arenacea Formation, FAA = Blue Clays Formation.


(a)
0 20 50 m
(b)
0 30 60 m
(c)
0 100 200 m
(d)
0 30 60 m
(e)
0 20 40 m
(f)
0 60 120 m

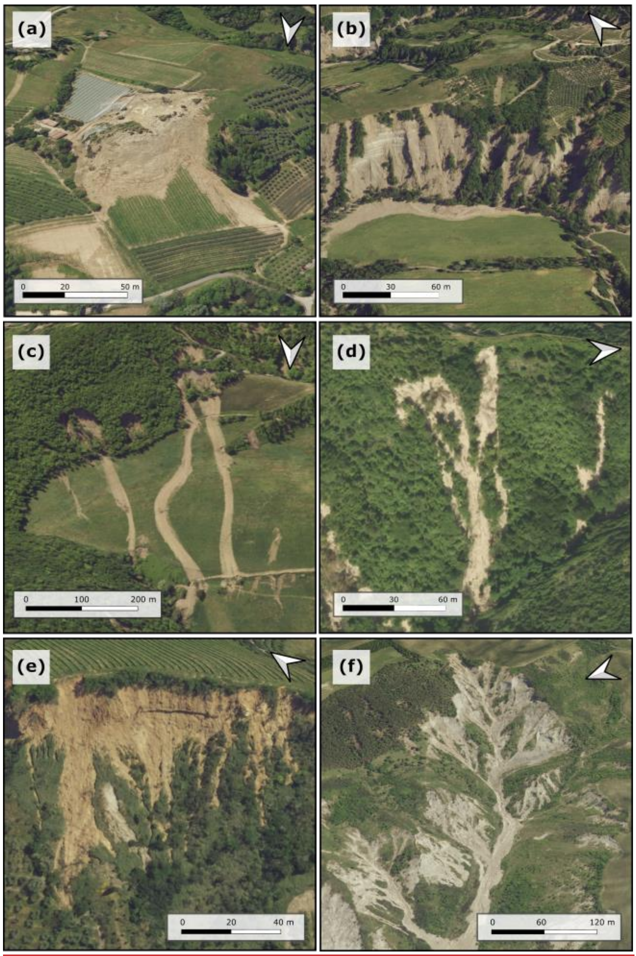


Figure 3. Examples of Landslides in Romagna from May 2023 event following Berti et al. 2025 classification.

a) Rockslide (RS) at Baccagnano, Brisighella, triggered during the night of May 2-3. (WGS84: 44.2069085, 11.7739792).

b) Debris slide (DS1) at Monte Albano, Casola Valsenio. (WGS84: 44.2356093, 11.6682287).

c) Debris flow with extended runout on gentle unforested slope (DF1) at Casola Valsenio. (WGS84: 44.20834, 11.64312).

d) Debris flow with limited runout on steep forested slope (DF2) at Modigliana. (WGS84: 44.15124, 11.75511).

e) Earth Slide (ES) at Brisighella. (WGS84: 44.23912, 11.78393).

f) Earth Flow (EF) at Predappio. (WGS84: 44.08947, 11.99627).

Images ~~sourced from the~~ © Regione Emilia-Romagna, Geoportale 3D ~~(Geoportale 3D Regione Emilia-Romagna~~https://mappe.regione.emilia-romagna.it~~.~~

ha formattato: Italiano (Italia)

~~3~~)

4. Methods

~~3~~4.1 Data

Deep-learning models for landslide mapping typically utilize a multi-layer input approach, integrating a combination of satellite or aerial imagery, topographical data, and land use information to capture the complex characteristics of landslides (Ghorbanzadeh et al., 2019, Meena et al., 2022). In our analysis, we considered the following seven layers (Fig. 4):

- L1) Post-event Sentinel-2 images: Four-~~bands~~band (RGB+NIR) satellite images ~~with a~~at 10 m spatial resolution, ~~captured~~acquired after the second rainfall event on May 23, 2023. Sentinel-2 Level-2A (L2A) products (bottom-of-atmosphere surface reflectance) were used. They represent the first post-event Sentinel-2 images obtained with minimal cloud cover (Fig. 4b).
- L2) Pre-event Sentinel-2 images: Four-~~bands~~band (RGB+NIR) satellite images ~~with a~~at 10 m spatial resolution ~~captured~~acquired in May 2022, one year prior to the event. Sentinel-2 Level-2A (L2A) products (bottom-of-atmosphere surface reflectance) were used. These images provide a similar vegetation state to that of May 2023~~,~~ and were ~~chosen to represent pre-event conditions~~selected due to the lack of cloud-free images in the two months preceding the event (Fig. 4a).
- L3) NDVI change Sentinel-2 map: Single-band raster generated by subtracting the pre-event NDVI map (Kriegler et al., 1969) derived from L2 from the post-event NDVI map created using L1 (Fig. 4c).
- L4) Post-event CGR images: Four bands (RGB+NIR) aerial photographs with a very high resolution of 0.2 m captured on May 23, 2023, shortly after the event. These images were committed by the Emilia-Romagna Region specifically for the disaster (Fig. 4f).
- L5) Pre-event AGEA images: Four bands (RGB+NIR) aerial photographs with a very high resolution of 0.2 m captured 3 years before the event (April to June 2020). These images were committed by the Agency for Agricultural Payments mainly for agricultural purposes (Fig. 4e).
- L6) NDVI change CGR map: Single-band raster generated by subtracting the pre-event NDVI map derived from L5 from the post-event NDVI map created using L4 (Fig. 4g).
- L7) Slope map: Single-band raster derived from the 5x5 m Digital Terrain Model (DTM) of the Emilia Romagna Region. The DTM, which is based on the 1:5000 scale Regional Technical Map, captures the pre-event slope morphology (Fig. 4d).

These layers were ~~selected~~combined to ~~simulate the diverse~~ define seven data-availability scenarios ~~that might be encountered during an emergency, based on the available data.~~ (cases 1-7 in Table ~~2 details the potential combinations of the seven layers, with~~ 3), each ~~combination corresponding to different~~ reflecting realistic operational conditions following a landslide-triggering event.

In all scenarios~~ explored in the study. In every case listed in Table 2, we assumed~~, the availability of a slope map (L7~~), as numerous global DEMs such as SRTM (Farr et al., 2007; NASA SRTM, 2013), ASTER (NASA Earthdata Search, 2023), ALOS (Rosenqvist et al., 2007; ASF DAAC, 2023),~~) was assumed. In the present study, L7 was derived exclusively from the regional 5 × 5 m DTM. However, this assumption reflects a more general operational condition, as slope information can be readily derived from freely available global digital elevation models (e.g., SRTM, ASTER, ALOS, Copernicus DEM), which are commonly accessible even in data-scarce emergency contexts.

Case 1 (Table 3) represents the most data-limited scenario, where only post-event Sentinel-2 imagery (L1) is available. Owing to its global coverage, short revisit time, and ~~Copernicus DEM (ESA, 2019) are readily accessible for~~ free accessibility, Sentinel-2 data often constitute the primary information source immediately after large-scale disasters (Wasowski et al., 2014; Yang et al., 2019; Ban et al., 2020).~~.~~

~~Often, Sentinel-2 images (L1) might be the only data accessible following a large-scale disaster. The Sentinel-2 satellites, part of the European Copernicus program, revisit any point on Earth every 5 days and provide free 10 m resolution images in the visible and near-infrared spectrum, which are vital for various applications including disaster management (Wasowki et al.,~~Case 2 considers situations in which pre-event Sentinel-2 imagery (L2) is also accessible, allowing the computation of an NDVI change map (L3). While this condition is frequently met, prolonged cloud cover or strong seasonal vegetation variability may prevent the availability of suitable pre-event images.

Cases 3 and 4 describe scenarios in which high-resolution post-event imagery is acquired, either alone (case 3) or in combination with Sentinel-2 data (case 4). Such data may originate from very-high-resolution satellite systems (e.g., WorldView, Pleiades, SkySat) or dedicated aerial surveys and substantially improve the detection of small or geomorphically subtle landslides.

Cases 5 and 6 address less common situations in which high-resolution imagery is available both before and after the event, optionally complemented by a high-resolution NDVI change map. The effectiveness of these scenarios depends strongly on the temporal proximity of the pre-event imagery to the triggering event.

Finally, case 7 represents the most data-rich configuration, combining all available layers (L1-L7).

~~2014; Yang et al., 2019; Ban et al., 2020). Numerous studies have utilized these open-access images to develop and evaluate automated methods for landslide mapping (Liu et al., 2021; Ghorbanzadeh et al., 2022; Lu et al., 2023; Notti et al., 2023). Consequently, L1 can be considered the basic data layer available in the aftermath of an event (case 1 in Table 2).~~

~~Frequently, though not always, it is possible to access pre-event Sentinel-2 images (L2) and compute the NDVI change map (L3) to improve the identification of landslides triggered by an event. However, these resources may not be available if the disaster was preceded by extended periods of cloudy weather during which significant changes in vegetation state occurred. In our study, we utilized cloud-free Sentinel-2 images taken during the same period the previous year, but this approach isn't always feasible in regions with intensive agricultural activity, where the state of vegetation can vary markedly from one year to the next. When pre-event Sentinel-2 images are available, automated methods can be trained using L1, L2, and L3 (case 2).~~

~~In developed countries, or when financial resources permit, high-resolution images can be obtained soon after the disaster. These images may come from satellite systems such as WorldView-3 and 4 (Maxar Technologies, 0.31 m), GeoEye-1 (Maxar Technologies, 0.41 m), Pleiades 1A and 1B (Airbus, 0.5 m), IKONOS (Maxar Technologies, 0.82 m), or SkySat (Planet Labs, 0.9 m), or they can be captured through dedicated aerial campaigns. The use of high-resolution imagery greatly enhances the precision of landslide mapping as it allows for the detection of even the smallest features and the clear identification of subtle geomorphic signs of instability. When high-resolution post-event images are available, they can be used on their own (case 3), or in combination with Sentinel-2 images (case 4), facilitating the capture of both small-scale and large-scale landslide features.~~

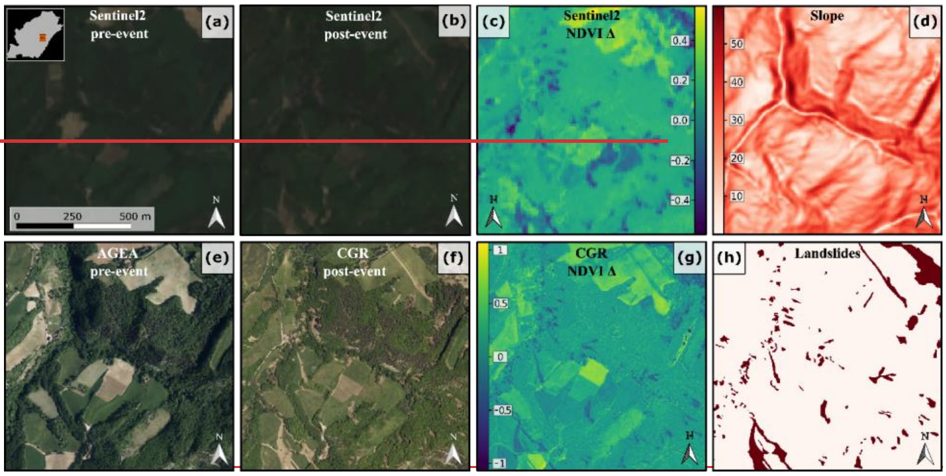
~~Less commonly, high-resolution images suitable for analysis are available before the event. The utility of these images' hinges on their timing relative to the disaster. Ideally, pre-event high-resolution images should capture the area immediately before the disaster to accurately pinpoint landslides triggered by the event and minimize false positives resulting from agricultural or natural vegetation changes. This requires appropriate satellite coverage of the area prior to the disaster. In our case, high-resolution aerial images were obtained three years before the event (L5); however, the general land cover remained unchanged, and no significant landslides had occurred in the area since then. Automated methods can be applied by integrating both pre- and post-event high-resolution images (case 5) or by also incorporating an NDVI change map derived from these datasets (case 6). A final case considers the combination of all available layers L1-7 (case 7).~~

Although the post-event (L4~~) and~~), pre-event (L5~~) aerial images, as well as the corresponding~~), and derived NDVI change ~~map~~ (L6~~),~~) aerial datasets were originally available at ~~a very high~~0.2 m spatial resolution~~ of 0.2 m~~, their use at ~~full resolution posed major computational issues. Specifically, attempts to process these layers at~~ native resolution ~~systematically~~ resulted in systematic out-of-memory ~~(OOM)~~ errors during ~~model~~ training and inference~~. This was~~ due to the large size of the input tensors~~ generated from 0.2 m imagery over wide areas, which exceeded the available GPU memory even when using optimized batch sizes and patching strategies.~~

. To ~~overcome this limitation and~~ ensure ~~stable execution of the deep learning pipeline, we resampled the L4, L5, and L6 layers to a coarser spatial resolution of 2 m. This down sampling allowed us to retain sufficient spatial detail for landslide identification, while significantly reducing the memory footprint of the input data. While the original 0.2 m resolution would have provided extremely fine detail, it exceeded the requirements~~ computational stability, were therefore downsampled to 2 m, representing a practical compromise between spatial detail and computational feasibility for ~~a~~ regional-scale ~~deep learning analysis. In~~mapping. By contrast, ~~the 2 m resolution still captures the essential geomorphological features relevant to landslide detection and represents a practical and effective compromise. This trade-off is consistent with the flexibility of deep learning models, which can learn robust spatial representations from moderate-resolution inputs when properly trained and validated~~Sentinel-2 layers (L1–L3) and the slope layer (L7) were resampled to the same 2 m grid solely to ensure pixel-wise alignment when used together with the aerial datasets; this step does not add information beyond their native resolution and was performed for operational consistency.

Table ~~2~~3. Overview of the various input layer configurations used during the training processes. 'NDVI' refers to the 'ΔNDVI-CGR' Change map created using CGR imagery, while 'ΔNDVI-S2' refers to the NDVI Change map created using Sentinel-2 imagery.

Name	**Model**	**Sentinel2 post** [L1]	**Sentinel2 pre** [L2]	**Sentinel2 ΔNDVI** [L3]	**CGR post** [L4]	**AGEA pre** [L5]	**CGR ΔNDVI** [L6]	**Slope** [L7]
U1	U-Net	✓	X	X	X	X	X	✓
S1	SegForm	✓	X	X	X	X	X	✓
U2	U-Net	✓	✓	✓	X	X	X	✓
S2	SegForm	✓	✓	✓	X	X	X	✓
U3	U-Net	X	X	X	✓	X	X	✓
S3	SegForm	X	X	X	✓	X	X	✓
U4	U-Net	✓	✓	✓	✓	X	X	✓
S4	SegForm	✓	✓	✓	✓	X	X	✓
U5	U-Net	X	X	X	✓	✓	X	✓
S5	SegForm	X	X	X	✓	✓	X	✓
U6	U-Net	X	X	X	✓	✓	✓	✓
S6	SegForm	X	X	X	✓	✓	✓	✓
U7	U-Net	✓	✓	✓	✓	✓	✓	✓
S7	SegForm	✓	✓	✓	✓	✓	✓	✓



Cases	Name	Model	Sentinel2 post [L1]	Sentinel2 pre [L2]	Sentinel2 ΔNDVI [L3]	CGR post [L4]	AGEA pre [L5]	CGR ΔNDVI [L6]	Slope [L7]
Case 1	U1	U-Net	✓	X	X	X	X	X	✓
	S1	SegForm	✓	X	X	X	X	X	✓
Case 2	U2	U-Net	✓	✓	✓	X	X	X	✓
	S2	SegForm	✓	✓	✓	X	X	X	✓
Case 3	U3	U-Net	X	X	X	✓	X	X	✓
	S3	SegForm	X	X	X	✓	X	X	✓
Case 4	U4	U-Net	✓	✓	✓	✓	X	X	✓
	S4	SegForm	✓	✓	✓	✓	X	X	✓
Case 5	U5	U-Net	X	X	X	✓	✓	X	✓
	S5	SegForm	X	X	X	✓	✓	X	✓
Case 6	U6	U-Net	X	X	X	✓	✓	✓	✓
	S6	SegForm	X	X	X	✓	✓	✓	✓
Case 7	U7	U-Net	✓	✓	✓	✓	✓	✓	✓
	S7	SegForm	✓	✓	✓	✓	✓	✓	✓

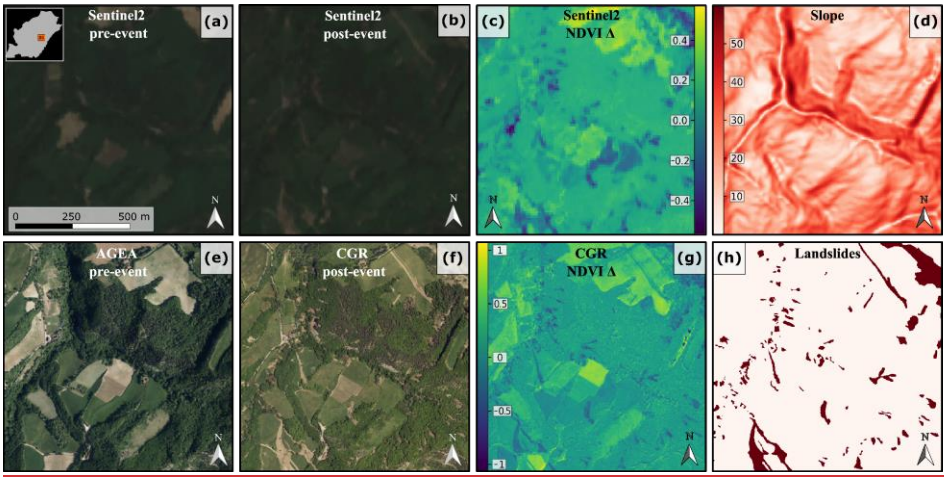


Figure 4. Example of the layers included in case 7, Tile No. 86. (**a**) Pre-event Sentinel-2 image (2022); (**b**) Post-event Sentinel-2 image; (**c**) NDVI change map derived from Sentinel-2 imagery; (**d**) Slope map; (**e**) AGEA imagery (2020), pre-event; (**f**) CGR aerial image (2023), post-event; (**g**) NDVI change map derived from high resolution imagery; (**h**) Binary raster showing manually mapped landslides (ground truth).

ha formattato: Tipo di carattere: Times New Roman, 12 pt, Non Corsivo, Colore carattere: Testo 1

Formattato: Normale, Allineato a sinistra, SpazioPrima: 0 pt, Dopo: 0 pt, Interlinea: singola

~~3.4 Deep Learning Semantic Segmentation Models~~

~~U-Net~~

~~U-Net is a convolutional neural network architecture originally designed for biomedical image segmentation by Ronneberger et al. (2015). It features an encoder-decoder structure where the encoder progressively reduces the spatial dimensions through convolutional and max-pooling layers, capturing contextual and high-level features. The decoder then up samples the features using transposed convolutions and concatenates them with corresponding encoder features through skip connections, which help retain spatial information and improve localization accuracy. This combination of context and localization has proven highly effective for segmentation tasks across various domains. U-Net's effectiveness in landslide mapping has been demonstrated in recent studies (Meena et al. 2021, 2022; Ghorbanzadeh et al. 2022, 2023; Nava et al. 2022).~~

~~In our study, we implemented the U-Net architecture manually in Python. The input shape of the images was defined according to the specific dimensions of the data used, ensuring the model could effectively process the input imagery. We set the number of scales to 2, which determines the number of downsampling steps in the encoder, and each scale included 2 convolutional layers, which helped capture intricate details and patterns within the data. The initial number of convolutional filters was set to 64. This parameter defines the number of filters applied in the first convolutional layer, which influences the model's ability to learn from the input data. As the network progresses through the layers, the number of filters increases, allowing the model to capture more complex features.~~

~~Our model was designed to output 2 classes: landslide and non-landslide. This binary classification approach enabled us to effectively differentiate between affected and unaffected areas in the imagery. The model was compiled using the Adam optimizer (Kingma and Ba, 2017), a learning rate of 0.0001 and a Dice loss function for 300 epochs. The early stopping mechanism was introduced with a patience of 20 epochs to prevent overfitting, based on validation loss monitoring.~~

~~All analyses were conducted using Python 3.8.18, TensorFlow 2.5.0, and CUDA 11.2.~~

~~*SegFormer*~~

~~SegFormer is a state-of-the-art architecture for semantic segmentation of multispectral landslide imagery (Xie et al., 2021). SegFormer, is designed to overcome limitations of previous vision Transformer models, such as their high computational cost, reliance on fixed input resolutions due to positional encodings, and inefficient multi-scale feature aggregation (Dosovitskiy et al., 2020). While the original SegFormer architecture was designed for 3-channel RGB images, the SegformerForSemanticSegmentation implementation from Hugging Face Transformers library (Wolf et al., 2020) allows for flexible input channel configuration. We leveraged this feature to adapt SegFormer to our multi channels imagery.~~

~~The SegFormer architecture consists of two main components (Xie et al., 2021):~~

1. ~~Hierarchical Transformer Encoder: This encoder employs a novel hierarchical structure with progressively larger stride and patch sizes, allowing the model to capture both fine and coarse features efficiently. It uses an efficient self-attention mechanism without positional encoding and incorporates a Mix-FFN block that includes a depth-wise convolution to enhance local spatial information aggregation.~~
2. ~~Lightweight All-MLP Decoder: Unlike complex decoder designs in previous models, SegFormer uses a simple multi-layer perceptron (MLP) decoder. This decoder fuses multi-level features from the encoder through a unified set of MLP layers, achieving per-pixel prediction without the need for complex operations like FPN or dilated convolutions.~~

~~SegFormer's design eliminates the need for positional encoding, allowing it to handle variable input resolutions without interpolation. This feature is particularly beneficial for our landslide imagery, which can vary in resolution and scale.~~

~~Our model was configured as the variant MiT-b0 (Xie et al., 2021) with the following parameters:~~

- ~~Hidden sizes: [32, 64, 160, 256]~~
- ~~Encoder depths: [2, 2, 2, 2]~~
- ~~Decoder hidden size: 256~~

~~The model was trained using Cross-Entropy Loss as the criterion. We employed the Adam optimizer with a learning rate of 0.001 for 300 epochs. The early stopping mechanism was introduced with a patience of 20 epochs to prevent overfitting, based on validation loss monitoring.~~

~~All analyses were conducted using Python 3.8.10, PyTorch (Paszke et al., 2019) 1.9.0, and CUDA 11.1.~~

~~3.5 Training-Testing Split~~

~~While~~ 4.4 Data preparation and training-testing design

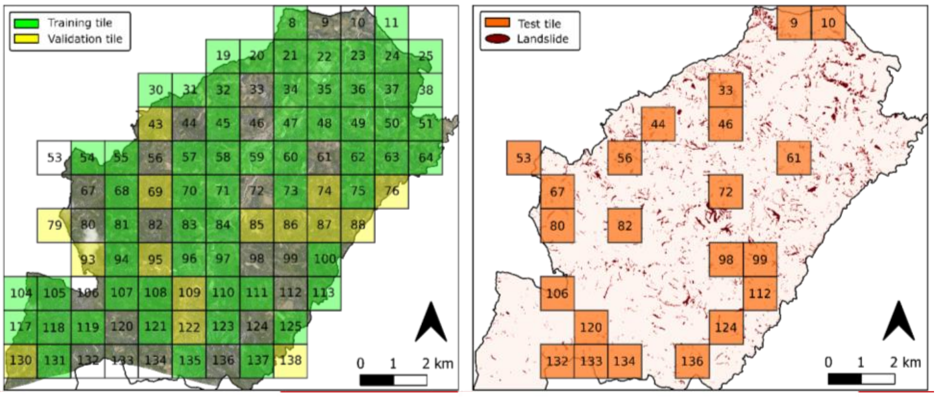
To reproduce operational conditions typical of post-event emergency mapping, the deep-learning workflow was designed to prioritise spatial generalisation rather than random pixel-level splitting. All input layers described in Section 3.1 were harmonised to a common spatial resolution and grid prior to training.

The training area (Casola Valsenio municipality) was subdivided into regular 1 × 1 km tiles (Fig. 5) to define consistent spatial sampling units and to reduce spatial autocorrelation effects. When resampled to a spatial resolution of 2 m, each tile corresponds to 512 × 512 pixels in the CGR imagery. Individual tiles typically include a large number of mapped landslides (often exceeding 50), providing a robust representation of landslide morphology and surrounding land cover.

Instead of conventional ~~practice often dictates a~~random splits (e.g. 70/30 or 80/20 ~~ratio for training and testing datasets (~~; Hastie et al., 2009), we adopted a ~~different approach~~strategy tailored to ~~simulate real-world~~ emergency ~~mapping conditions.~~response scenarios. The ~~model was~~models were trained exclusively ~~trained~~ on data from ~~the~~ Casola Valsenio ~~municipality~~ and subsequently applied, without retraining, to three ~~additional~~neighbouring municipalities, Predappio, Modigliana, and Brisighella, for independent testing. This setup simulates a realistic ~~emergency response scenario where; after mapping~~operational context in which a ~~relatively small area, the~~ model trained on a limited reference area is rapidly deployed to ~~rapidly~~ map other affected ~~areas, relying solely on prior training~~regions.

~~To facilitate analysis, the train area was segmented into 1x1 km grid tiles, as illustrated in Fig. 5. Each 1 km² tile corresponds to 512x512 pixels in the CGR imagery when resampled at 2m resolution. Generally, a single grid tile in these areas encompasses over 50 landslides, providing a robust representation of landslide patterns.~~

~~For model training, we adopted a stratified random sampling approach. Out of the 97 total tiles in Casola Valsenio, 62% (60 tiles) were randomly allocated for training, 15% (15 tiles) for validation, and the remaining 23% (22 tiles) were reserved for initial testing and further validation (Fig. 5). This division was designed to maximize the training set while ensuring sufficiently large and reliable validation and test sets, given the need to apply the model to other municipalities.~~



Within Casola Valsenio, a stratified random sampling strategy was used to partition the tiles into training, validation, and internal testing subsets. Out of 97 tiles, 60 (62%) were assigned to training, 15 (15%) to validation, and 22 (23%) to internal testing (Fig. 5). External evaluation was then conducted on all tiles from the three remaining municipalities, providing a spatial hold-out test to assess cross-municipality generalisation. This tile-based spatial splitting yields a strongly imbalanced pixel-level segmentation problem, because landslide pixels represent only a small fraction of the mapped area. In Casola Valsenio (training area), landslide pixels account for ~ 4% of the total pixels (non-landslide:landslide ratio ~24:1). A comparable imbalance characterises the external test areas, with ~6% landslide pixels in Predappio (~ 16:1), ~7.6% in Modigliana (~12:1), and ~5% in Brisighella FMA (~19:1). The most extreme imbalance occurs in Brisighella FAA, where landslide pixels represent ~2% of the pixels (~48:1). The implications of this class imbalance for model evaluation are discussed in Section 4.7.

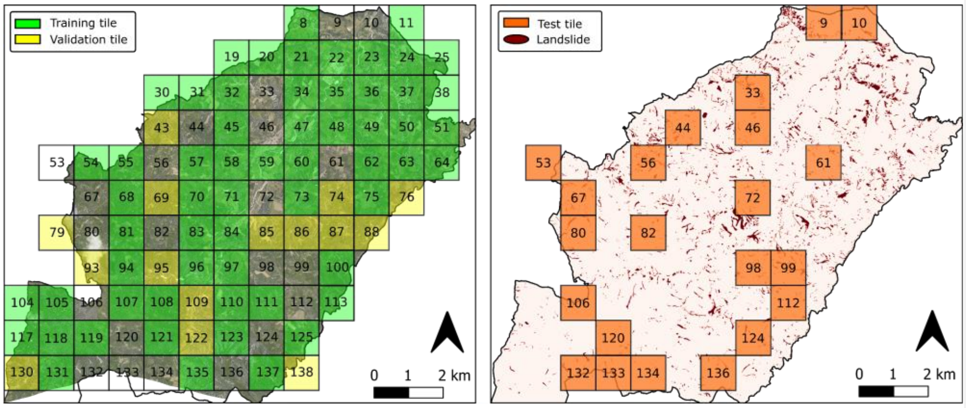


Figure 5. Training and test tiles used for models' construction in Casola Valsenio, illustrating the spatial split of the area.

~~3.6 Model Application~~

4.5 Deep Learning Semantic Segmentation Models

To evaluate the performance of different deep-learning paradigms under variable data-availability scenarios (Table 3), two semantic segmentation architectures were implemented: a convolutional encoder-decoder network (U-Net) and a Transformer-based segmentation model (SegFormer). Schematic representations of both architectures are provided in Fig. 6 to facilitate understanding and comparison.

4.5.1 U-Net

U-Net is a convolutional neural network originally developed for biomedical image segmentation (Ronneberger et al., 2015) and subsequently adopted in a wide range of geomorphological and Earth-observation applications. Its encoder-decoder structure enables the extraction of contextual information while preserving fine-scale spatial detail, a key requirement for accurate landslide delineation. The suitability of U-Net for landslide mapping has been demonstrated in several recent studies (Meena et al., 2021, 2022; Ghorbanzadeh et al., 2022, 2023; Nava et al., 2022).

ha formattato: Tipo di carattere: 11 pt

Formattato: Allineato a sinistra, SpazioPrima: 0 pt, Dopo: 8 pt, Interlinea: multipla 1,08 ri

The architecture consists of a contracting encoder path and an expanding decoder path. In the encoder, spatial resolution is progressively reduced through convolutional and max-pooling layers to capture high-level contextual features. In the decoder, feature maps are upsampled using transposed convolutions and concatenated with corresponding encoder features via skip connections, which preserve spatial information and improve boundary localisation (Fig. 6a).

In this study, U-Net was implemented manually in Python and configured for binary semantic segmentation (landslide vs non-landslide). The network employed two downsampling scales, each comprising two convolutional layers. The initial number of filters was set to 64 and increased with network depth to capture progressively more complex features.

Training was performed using the Adam optimiser (Kingma and Ba, 2017) with a learning rate of 1×10^{-4} and Dice loss, for up to 300 epochs. Early stopping with a patience of 20 epochs was applied based on validation loss. All experiments were conducted using Python 3.8.18, TensorFlow 2.5.0, and CUDA 11.2.

4.5.2 SegFormer

SegFormer is a Transformer-based architecture specifically designed for efficient semantic segmentation of high-resolution and multispectral imagery (Xie et al., 2021). It addresses several limitations of earlier Vision Transformer models, including high computational cost, dependence on positional encodings, and inefficient multi-scale feature aggregation (Dosovitskiy et al., 2020).

The SegFormer architecture comprises two main components (Xie et al., 2021): (i) a hierarchical Transformer encoder that captures features at multiple spatial scales using efficient self-attention without positional encodings and Mix-FFN blocks with depth-wise convolutions, and (ii) a lightweight all-MLP decoder that fuses multi-level encoder features for dense per-pixel prediction (Fig. 6b). The absence of positional encodings allows the model to handle variable input resolutions without interpolation, which is advantageous for heterogeneous landslide imagery.

We adopted the MiT-B0 variant (Xie et al., 2021), with hidden sizes [32, 64, 160, 256], encoder depths [2, 2, 2, 2], and a decoder hidden size of 256. Although SegFormer was originally designed for three-channel RGB imagery, we exploited the flexibility of the SegformerForSemanticSegmentation implementation provided by the Hugging Face Transformers library (Wolf et al., 2020) to accommodate multi-channel inputs.

Training was performed using Cross-Entropy loss and the Adam optimiser with a learning rate of 1×10^{-3}, for up to 300 epochs, with early stopping based on validation loss (patience = 20 epochs). Computations were carried out using Python 3.8.10, PyTorch 1.9.0 (Paszke et al., 2019), and CUDA 11.1.

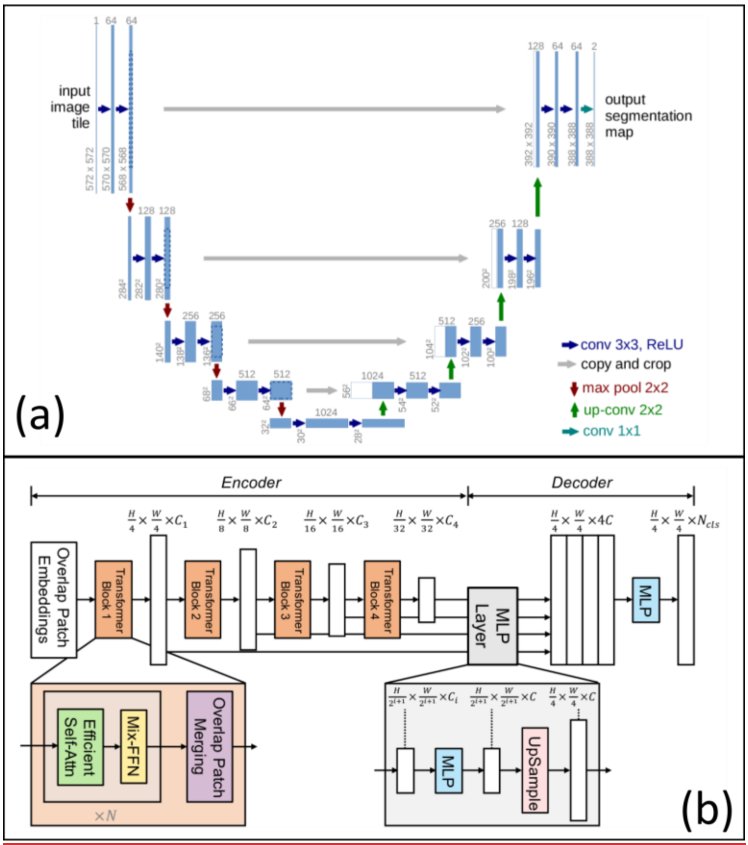


Figure 6. Conceptual overview of the deep-learning semantic segmentation architectures used in this study. (**a**) U-Net encoder-decoder architecture with skip connections, illustrating the contracting path for contextual feature extraction and the expanding path for precise spatial localisation (adapted from Ronneberger et al., 2015). (**b**) SegFormer architecture, composed of a hierarchical Transformer encoder (MiT) with overlapping patch embeddings and efficient self-attention, and a lightweight all-MLP decoder for multi-scale feature fusion and dense per-pixel prediction (adapted from Xie et al., 2021).

4.6 Model application and geological transferability

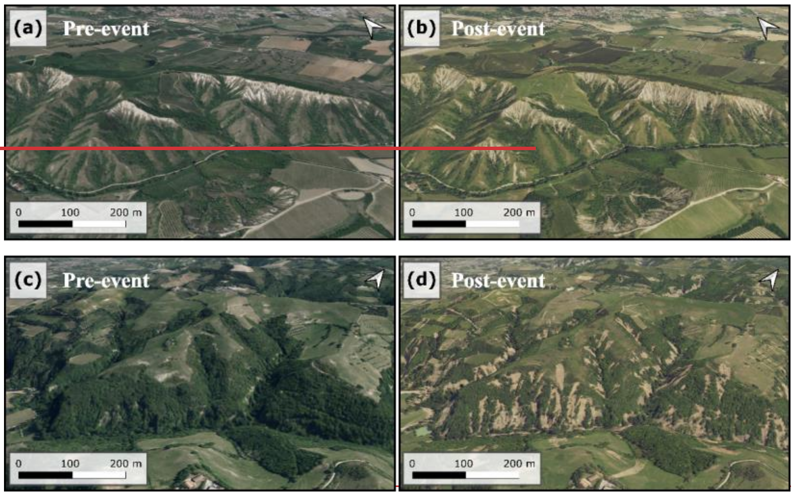
Following ~~the~~ training ~~phase, we applied,~~ the models were applied to the municipalities of Predappio, Modigliana, and Brisighella. ~~Crucially, while~~ Although ground-~~-~~truth ~~(Fig. 1a) data was~~ landslide inventories were available for these ~~municipalities, we deliberately chose~~ areas (Fig. 1a), they were not ~~to use it for model~~ used during training~~. Instead, this data was~~ and were reserved exclusively for ~~evaluating the performance of our models.~~ model evaluation.

A ~~significant constraint~~key limitation of the ~~analysis arises from~~training dataset is the predominance of the Marnoso-Arenacea Formation (FMA) in ~~the training dataset from~~ Casola Valsenio (Fig. 3a). To explicitly assess ~~the models' ability to map landslides in varied~~ geological ~~contexts, we divided~~tralnsferability, the Brisighella municipality was subdivided into two distinct ~~geological zones~~lithological domains: one dominated by the Marnoso-Arenacea Formation (Brisighella FMA) and ~~another characterized by the predominance of fine-grained rock from the~~ one characterised by Pliocene Blue Clays (Brisighella FAA~~)~~ (; Fig. 3d).~~ This latter area~~

The FAA domain is ~~characterized~~dominated by earth flows (EF) and earth slides (ES), ~~which exhibit fundamentally different morphological, and soil cover features compared to landslides present in the Marnoso-Arenacea Formation. Figure 6 shows representative examples of these contrasting landslide types. The FAA zone, underlain by the Pliocene Blue Clays, is typified by slow-moving earth flows (EF) and earth slides (ES), which~~ typically developed within fine-grained, clay-rich materials and often ~~occur within~~ associated with badland-like morphologies. ~~These formations exhibit high plasticity and low permeability, conditions that favor surface runoff and shallow erosion rather than deep infiltration. As a result, slope instability is driven primarily by progressive degradation and surface washout, leading to broad~~These landslides generally exhibit elongated shapes, lobate ~~flow patterns rather than discrete failures. The vegetative cover is often~~toes, and sparse or degraded~~, further intensifying surface erosion during rainfall events.~~

vegetation cover, reflecting repeated reactivation processes (Fig. 7a-b). In contrast, landslides in the FMA ~~zone~~domain, both in Brisighella and in the training area~~ of Casola Valsenio~~, are ~~dominated by~~mainly debris slides (DS) and debris flows (DF~~). These failures typically originate from~~) triggered within thin~~, unconsolidated~~ colluvial layers ~~that overlie the flysch bedrock, and are triggered by intense rainfall infiltration. The resulting movements are generally more sudden~~overlying flysch bedrock. These failures tend to be rapid, with well-defined ~~scarp~~scarps and runout zones, and commonly affect ~~areas with denser vegetation cover.~~densely vegetated slopes (Fig. 7c-d).

~~By explicitly separating these two lithological domains within Brisighella, we aimed to evaluate whether a model trained solely on FMA-type landslides could generalize to areas with fundamentally different geological and geomorphological settings. This setup also allowed us to assess the sensitivity of the model to lithology-driven differences in landslide type, geometry, and spatial pattern—an essential step toward developing mapping tools that are robust across variable terrain conditions.~~



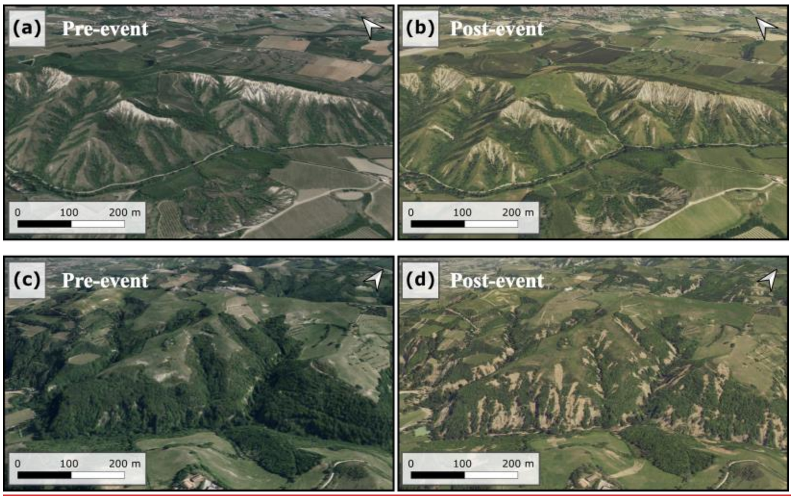


Figure ~~6~~7. Comparison examples are shown between pre-event (**a**) and post-event (**b**) earth flow (EF) landslides in the Blue Clays Formation (FAA) and pre-event (**c**) and post-event (**d**) debris slide (DS) in the Marnoso-Arenacea Formation (FMA), both triggered by the May 2023 disaster in Brisighella.

Formattato: SpazioPrima: 0 pt, Dopo: 0 pt

~~The final evaluation of model generalization was conducted on the entire set of grid tiles from the three remaining municipalities: 90 tiles from Predappio, 105 from Modigliana, and 196 from Brisighella. As such, although the core training-validation-testing split within Casola Valsenio followed a 62/15/23 ratio, the effective test set includes not only the 22 held-out tiles from Casola Valsenio, but also 100% of the 413 tiles from the other three municipalities, a spatial hold-out test designed to assess cross-municipality generalization.~~

~~**Table 3.** Overview of the utilization of 1 km² tiles across the analyzed municipalities, illustrating how each area was segmented for model training, validation, testing and prediction purposes.~~

~~Area~~	~~Tiles [1 km²]~~	~~Use~~
Casola Valsenio	97	60 train / 15 validation / 22 test
Brisighella FMA	116	Prediction
Brisighella FAA	80	Prediction
Modigliana	105	Prediction
Predappio	90	Prediction

~~3~~Images © Regione Emilia-Romagna, Geoportale 3D (https://mappe.regione.emilia-romagna.it)

By explicitly separating these lithological contexts, we evaluated whether models trained solely on FMA-type landslides could generalise to areas characterised by fundamentally different landslide processes, geometries, and surface expressions. A summary of the spatial extent of each lithological unit within the test municipalities is provided in Table 4 to support interpretation of the evaluation results.

Table 4. Overview of the 1 km² tiles used for model training, validation, and testing across the analysed municipalities. The table reports the number of tiles assigned to each dataset split and their distribution within the main lithological units, namely the Marnoso-Arenacea Formation (FMA) and the Pliocene Blue Clays (FAA). External testing was performed on 100% of the tiles from Predappio, Modigliana, and Brisighella to assess cross-municipality and cross-lithology model generalisation.

Municipality	Total tiles [1 km^2]	Training	Validation	Internal test	External test	FMA 'Unit 7'	FAA 'Unit 1'
Casola Valsenio	97	60	15	22	–	97	0
Predappio	90	–	–	–	90	75	15
Modigliana	105	–	–	–	105	99	6
Brisighella FMA	116	–	–	–	116	116	0
Brisighella FAA	80	–	–	–	80	0	80

4.7 Evaluation Metrics and Expert Judgment

4.7.1 Quantitative Metrics

Formattato: Default

To assess the performance of our landslide detection models, we employed two widely used metrics in semantic segmentation tasks: the F1 score and the Intersection over Union (IoU). These metrics were calculated on the entirety of the test images, rather than averaging scores across individual cells, to provide a comprehensive evaluation of the models' performance.

The F1 score~~, also known as the ~~ (Dice coefficient~~,~~) is the harmonic mean of precision and recall (Tharwat, 2020). ~~It~~We report F1 alongside IoU because the pixel-wise segmentation task is ~~particularly useful in~~strongly imbalanced ~~classification problems, such as~~(Section 4.4), with landslide ~~detection, where~~pixels representing only a small fraction of the ~~positive class (landslide) is typically much smaller than the negative class (non-landslide).~~mapped area. The F1 score is calculated as:

$$\text{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{1}$$

where Precision = TP / (TP + FP) and Recall = TP / (TP + FN), with TP, FP, and FN representing True Positives, False Positives, and False Negatives, respectively.

The Intersection over Union (IoU), also known as the Jaccard index, is another common metric for assessing segmentation accuracy (Rezatofighi et al., 2019). It measures the overlap between the predicted segmentation mask and the ground truth mask. The IoU is calculated as:

$$\text{IoU} = \frac{TP}{\text{TP} + \text{FP} + \text{FN}} \tag{2}$$

Both metrics range from 0 to 1, with 1 indicating perfect prediction. By using these metrics on the entire test images, we can evaluate how well our models perform in detecting and delineating landslides across varied landscapes and geological contexts.

4.7.2 Expert Judgment

In automated landslide mapping, a higher algorithmic score does not necessarily correspond to a higher quality map (Zhang et al., 2018; Isensee et al., 2021). This issue is particularly relevant in automated landslide mapping, where the practical utility of the map often depends more on its interpretability and coherence with geomorphological reality than on its statistical performance alone.

For example, a model might overpredict large landslide areas to optimize pixel-wise metrics like IoU, while neglecting smaller or fragmented features that are critical for field validation and emergency response. Conversely, models may introduce noise or irregular boundaries that artificially increase recall but reduce the practical readability of the map.

While this issue is typically addressed by selecting evaluation metrics that emphasize specific components of the confusion matrix (e.g., recall over precision), in emergency scenarios not all components carry the same operational weight. For instance, in crisis management, a higher recall may be preferred to avoid missing active landslides, even if this comes at the cost of more false positives. In contrast, in densely urbanized areas or during resource-constrained interventions, precision may become more important to avoid unnecessary alarms and misallocation of efforts.

To address this challenge, three of the authors in this study provided their expert judgment on the quality of the automatically generated landslide maps. This evaluation was conducted without prior knowledge of the scores obtained by these maps or the training input layers used to generate them.

The assessment focused on seven common error types frequently encountered in automated landslide mapping:

E1 ~~–~~- False positives in ~~agricultural~~plowed fields, where ~~cultivated~~plowed land is mistakenly classified as landslide;

E2 ~~–~~- False positives along roads and in urban areas, where the absence of vegetation may be misinterpreted as landslide activity;

E3 ~~–~~- False positives along riverbeds, often due to confusion with recent sediment scouring or bedload deposits;

E4 ~~–~~- False negatives in areas with known landslides;

E5 ~~–~~- Over segmentation, referring to the fragmentation of a single landslide into multiple small polygons;

E6 ~~–~~- Inaccurate delineation of landslide boundaries;

E7 ~~–~~- Omission errors in shadowed areas, such as steep, shaded slopes or regions obscured by cloud cover.

For each error type, the experts assigned a score of 0 (limited error), 1 (moderate error), or 2 (relevant error) to indicate the severity of the issue. These scores were then summed and normalized to calculate an Expert-based Performance Index (EPI):

$$EPI = \frac{14 - \sum_{i=1}^{i=7} E_i}{14}$$

The index ranges from 0 to 1, with 1 representing the highest mapping performance. This evaluation was independently carried out by each of the three experts for all 16 cases included in the analysis (see Table ~~2~~4). The experts assessed the model cases without prior knowledge of the specific configurations, thereby minimizing potential bias.

4.8 Identification of At-Risk Buildings

ha formattato: Tipo di carattere: Non Grassetto, Sottolineato, Legature: Standard + Contestuali

In the context of emergency mapping, a key objective was to identify buildings at risk due to the landslides triggered by the May 2023 event. This was a critical task for the Civil Protection as it allowed for timely intervention and damage assessment. Given the extensive spatial distribution of the damage caused by the event, it was essential to develop a systematic method to detect and assess the buildings at risk. Additionally, accurate damage estimation was necessary to request financial support under the European Union Solidarity Fund (EUSF), adhering to the EU's required timelines for funding requests.

To evaluate the effectiveness of our automated landslide mapping models in identifying at-risk buildings, we created a 20-meter buffer around the predicted landslide boundaries. This buffer was compared to the list of buildings that the Civil Protection had classified as "at risk," based on manual mapping (Pizziolo et al., 2024; Berti et al., 2025). The buildings identified by the automated models were then compared to these reference buildings, allowing us to assess the models' capacity for accurate identification. The confusion matrix was used to quantify the performance of the models in identifying buildings at risk, including both false positives (incorrectly flagged buildings) and false negatives (missed buildings).

This methodology ensured that the models were tested in the real-world context of emergency response, where timely and accurate damage assessments are crucial for effective resource allocation and intervention. The results of this methodology are discussed further in Section 6.3, while the quantitative outcomes of our models are presented in Section 5.3.

5. Results

~~4~~5.1 Model results and performance

The results of the analysis are ~~presented~~reported in Table ~~4. The table details~~5, which summarizes the F1 and IoU scores obtained by comparing the automated landslide maps ~~produced~~generated by the two ~~machine~~ deep-learning models (U~~=~~ = U-Net; S~~=~~ = SegFormer) ~~against~~with the manually ~~created maps. The rows list the seven different~~mapped reference inventory. Seven combinations of input layers~~, numbered from 1 to~~ (cases 1-7~~, to represent a~~; Table 1) were tested, representing progressively richer ~~dataset (refer to Table 1).~~input information. The models were~~ initially~~ trained using data from Casola Valsenio and subsequently applied to four ~~additional~~independent test areas ~~(Predappio, Modigliana,~~: Brisighella FMA, ~~and~~ Brisighella FAA~~). Overall, this combination of 7 input datasets, 4 areas, and 2 models resulted in~~, Modigliana and Predappio. In total, 56 automated landslide maps~~.~~ were produced (7 input configurations × 4 test areas × 2 models).

Across all municipalities, the highest performance was obtained by the most information-rich configurations, particularly U7 and U6 for U-Net and S7 for SegFormer. In Casola Valsenio, the best configuration reached F1 = 0.73 and IoU = 0.57 (U7), while in Brisighella FMA, Brisighella FAA, Modigliana and Predappio the highest F1-scores were consistently around 0.60–0.63 (e.g., U7 = 0.56–0.63; S7 = 0.60–0.61). Performance was systematically lower in Brisighella FAA, where the best configurations reached F1 = 0.53 and IoU = 0.36 (U6/U7), and F1 = 0.52 and IoU = 0.35 (S7) (Table 5).

Despite relying on a reduced set of inputs, the Sentinel-2-based configuration S2 (post-event Sentinel-2 plus Sentinel-2 NDVI change) achieved competitive results across municipalities. For instance, in Casola Valsenio S2 reached F1 = 0.62 and IoU = 0.45, compared with F1 = 0.73 and IoU = 0.57 for the best-performing configuration (U7). In Brisighella FMA, Modigliana and Predappio, S2 yielded F1-scores between 0.54 and 0.57 (IoU = 0.37–0.40), which are close to the values obtained by the strongest configurations (typically within ~0.03–0.06 in F1 and ~0.03–0.06 in IoU). In Brisighella FAA, S2 remained among the most stable configurations (F1 = 0.47; IoU = 0.31), whereas some intermediate cases showed marked drops (e.g., U3: F1 = 0.29; IoU = 0.17).

Overall, increasing the number of input layers resulted in measurable but limited performance improvements. The largest gain was associated with the inclusion of NDVI change information (case 2 vs. case 1), while subsequent additions of high-resolution inputs generally produced smaller incremental changes. These results indicate that streamlined configurations based on Sentinel-2 data (case 2), optionally complemented by high-resolution imagery, can achieve segmentation performance close to that obtained using the full set of available inputs (Table 5).

Table 5. F1-score and Intersection over Union (IoU) obtained by the two deep-learning semantic segmentation models (U = U-Net; S = SegFormer) for the seven input-layer configurations (cases 1–

7; Table 3), evaluated against the manually mapped reference inventory. Models were trained in Casola Valsenio and applied without retraining to the external test municipalities (Brisighella FMA, and Brisighella FAA, Modigliana, Predappio). The table therefore reports 56 model outputs (7 cases × 4 municipalities × 2 models).

Cases	Models' Name	Casola Valsenio		Brisighella FMA		Brisighella FAA		Modigliana		Predappio	
		F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
Case 1	U1	0.56	0.39	0.50	0.33	0.32	0.19	0.48	0.31	0.42	0.27
	S1	0.47	0.31	0.46	0.30	0.34	0.21	0.45	0.29	0.46	0.30
Case 2	U2	0.64	0.47	0.57	0.40	0.38	0.24	0.55	0.38	0.52	0.35
	S2	0.62	0.45	0.57	0.40	0.47	0.31	0.57	0.40	0.54	0.37
Case 3	U3	0.70	0.54	0.55	0.38	0.29	0.17	0.45	0.29	0.39	0.24
	S3	0.54	0.37	0.49	0.32	0.32	0.19	0.44	0.29	0.44	0.28
Case 4	U4	0.72	0.56	0.62	0.45	0.47	0.31	0.58	0.41	0.55	0.38
	S4	0.68	0.52	0.62	0.45	0.49	0.33	0.60	0.43	0.59	0.42
Case 5	U5	0.71	0.55	0.62	0.45	0.50	0.34	0.56	0.39	0.53	0.36
	S5	0.65	0.48	0.60	0.43	0.43	0.28	0.56	0.39	0.56	0.39
Case 6	U6	0.72	0.56	0.63	0.46	0.53	0.36	0.59	0.42	0.56	0.38
	S6	0.65	0.48	0.58	0.41	0.41	0.26	0.54	0.37	0.53	0.36
Case 7	U7	0.73	0.57	0.63	0.46	0.53	0.36	0.60	0.42	0.56	0.39
	S7	0.66	0.49	0.61	0.44	0.52	0.35	0.60	0.43	0.60	0.43

Figure 8 illustrates the evolution of F1-scores across the seven input configurations for each municipality. For both models, the inclusion of NDVI change information (case 2) led to an increase in F1-score relative to case 1. In Casola Valsenio, F1 increased from 0.56 to 0.70 for U-Net and from 0.47 to 0.54 for SegFormer. In Modigliana and Brisighella FAA, the corresponding increases were approximately +0.15 to +0.16 for U-Net. Beyond case 2, U-Net showed gradual increases in F1-score of 0.01-0.04 per configuration up to case 7, whereas SegFormer exhibited smaller variations, generally within ±0.02, indicating a lower sensitivity to further input enrichment.

Model performance varied across municipalities. In Casola Valsenio, U-Net F1-scores increased from 0.56 (U1) to 0.73 (U7), while SegFormer ranged from 0.47 (S1) to 0.66 (S7). In Predappio, Modigliana, and Brisighella FMA, U-Net F1-scores were between 0.55 and 0.63, whereas SegFormer values remained more stable, typically between 0.55 and 0.60. Brisighella FAA consistently showed the lowest performance for both models: U-Net F1-scores ranged from 0.29 (U3) to 0.53 (U7), and SegFormer values ranged from 0.32 to 0.52, with no systematic improvement beyond case 2. IoU values exhibited the same spatial pattern, decreasing from 0.35-0.46 in FMA-dominated areas to 0.17-0.36 in the FAA domain.

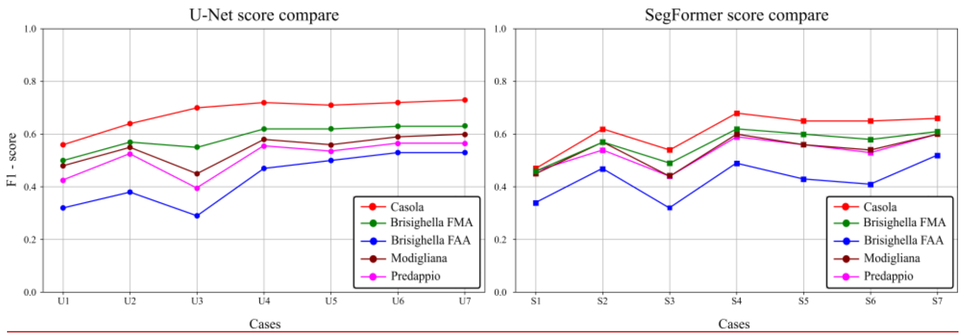


Figure 8. ~~In the tested configurations, the models U6, U7, and S7 achieved the highest F1 and IoU scores. These models incorporate nearly all (U6) or all (U7-S7) of the seven input layers, and clearly benefit from the high-resolution pre- and post-event images. However, the performance gains from utilizing all available layers are relatively modest. Notably, model S2, which employs only Sentinel-2 images and Sentinel-2 NDVI, achieves results close to those of the top-performing models, with differences in F1 and IoU scores frequently around a margin of ~ 0.01. Similar observations apply to models U4 and S4, which rely on Sentinel-2 products and post-event high-resolution images. This suggests that while increasing input information enhances model performance, even more streamlined configurations can still yield robust and competitive mapping outcomes. The good performance of U4 and S4 models highlights the effectiveness of integrating high-resolution imagery alongside Sentinel-2-derived indices, even when fewer spectral inputs are available.~~

~~**Table 4.** Quality of the automatic mapping, based on F1 and IoU scores, achieved by U-Net and SegFormer algorithms with different combinations of input layers, detailed in Table 2.~~

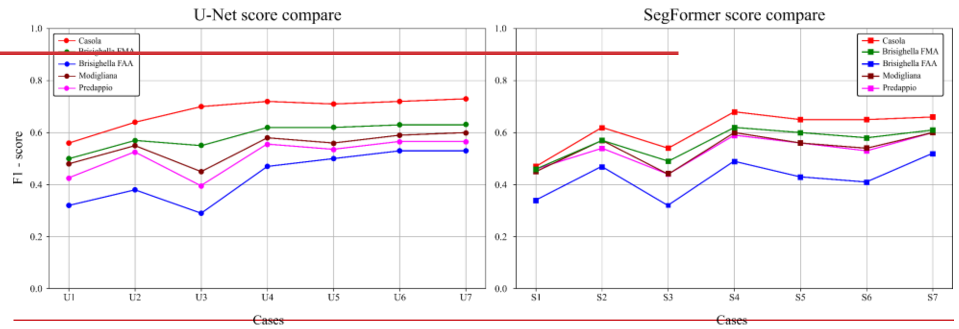
Models' Name	Casola Valsenio		Predappio		Modigliana		Brisighella FMA		Brisighella FAA	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
U1	0.56	0.39	0.42	0.27	0.48	0.31	0.50	0.33	0.32	0.19
S1	0.47	0.31	0.46	0.30	0.45	0.29	0.46	0.30	0.34	0.21
U2	0.64	0.47	0.52	0.35	0.55	0.38	0.57	0.40	0.38	0.24
~~S2~~	~~0.62~~	~~0.45~~	~~0.54~~	~~0.37~~	~~0.57~~	~~0.40~~	0.57	0.40	0.47	0.31
U3	0.70	0.54	0.39	0.24	0.45	0.29	0.55	0.38	0.29	0.17
S3	0.54	0.37	0.44	0.28	0.44	0.29	0.49	0.32	0.32	0.19
U4	0.72	0.56	0.55	0.38	0.58	0.41	0.62	0.45	0.47	0.31
S4	0.68	0.52	0.59	0.42	0.60	0.43	0.62	0.45	0.49	0.33
U5	0.71	0.55	0.53	0.36	0.56	0.39	0.62	0.45	0.50	0.34
S5	0.65	0.48	0.56	0.39	0.56	0.39	0.60	0.43	0.43	0.28
U6	0.72	0.56	0.56	0.38	0.59	0.42	0.63	0.46	0.53	0.36
S6	0.65	0.48	0.53	0.36	0.54	0.37	0.58	0.41	0.41	0.26
U7	0.73	0.57	0.56	0.39	0.60	0.42	0.63	0.46	0.53	0.36
S7	0.66	0.49	0.60	0.43	0.60	0.43	0.61	0.44	0.52	0.35

~~Figure 7 shows the F1-score trends across all seven input layer combinations (cases 1 to 7) for the five evaluated areas. In both models, a marked increase in performance is observed between case 1 and case 2, particularly in Casola Valsenio, Modigliana, and Brisighella FAA. This highlights the importance of including NDVI change information alongside Sentinel-2 imagery to reach acceptable segmentation performance. From case 2 onward, the models tend to stabilize, with U-Net showing a more pronounced benefit from additional input layers, especially in cases 4 to 7. SegFormer, by contrast, reaches near-plateau performance earlier and exhibits smaller gains as layers are added.~~

~~Performance across the different municipalities reveals further differences between the two models. U-Net shows a clear separation in F1-scores depending on the area: it performs best in Casola (reaching 0.73 in case 7), moderately well in Brisighella FMA, Modigliana, and Predappio, and significantly worse in Brisighella FAA (with 0.29 in case 3 and 0.53 in case 7). This trend reflects U-Net's sensitivity to geological conditions.~~

~~SegFormer, on the other hand, provides more uniform scores across areas, although generally lower than U-Net in Casola. Its performance ranges between 0.47–0.68 in Casola and remains around 0.55–0.60 in Predappio, Modigliana, and Brisighella FMA. Brisighella FAA again represents a challenge, with F1-scores fluctuating and even slightly declining from case 2 onward, suggesting that additional input layers do not consistently improve results in this more complex setting.~~

~~Overall, both models identify Brisighella FAA as the most difficult area, with the lowest F1-scores recorded. However, while U-Net appears to benefit more from richer input combinations, SegFormer demonstrates greater generalization ability across areas with different lithological and morphological characteristics.~~



~~**Figure 7.** F1-score comparison across all seven input layer combinations (cases 1–7) for each municipality, as obtained by the U-Net and SegFormer models.~~

~~As a~~ F1-score comparison across the seven input-layer configurations for U-Net (left) and SegFormer (right). The numbering matches the input-layer cases defined in Table 3 (e.g., U1 and S1 refer to Case 1; U2 and S2 refer to Case 2), and values are summarised in Table 5.

Figures 9 and 10 complement ~~to~~ the quantitative ~~evaluation, Figures 8 and 9 provide illustrative examples~~ results by providing a visual interpretation of model performance ~~in real-world conditions. These visual comparisons, focusing on the best-performing configurations (U-Net case 7 and SegFormer case 7), highlight both successful detections and common failure modes observed across the test municipalities.~~

~~Figure 8 displays cases where~~ under contrasting conditions. Figure 9 illustrates representative cases in which both models performed well, ~~accurately mapping landslides in~~showing a strong spatial agreement between the manually mapped inventory and the automated outputs. In these panels, the overlap between the reference landslide map (blue) and the predicted map (yellow) produces a beige colour, corresponding to true positives and indicating accurate landslide delineation. These examples mainly refer to debris slides and debris flows occurring in well-illuminated areas with widespread vegetation removal, ~~typical of debris slides (DS) and debris flows (DF). These landslide types, which together accounted for about 81% of all events in May 2023, were consistently recognized, especially in severely impacted zones where spectral and texture changes were prominent. Figure 9, by contrast, presents a selection of cases where both models encountered notable difficulties, leading to different types of misclassifications:~~where spectral and textural changes are pronounced.

- ~~Casola Valsenio: Despite being the training area, the models occasionally misinterpreted the spatial extent of landslide polygons. In some cases, large landslides are fragmented into several smaller segments, resulting in False Negatives and underestimation of the actual affected area. This issue is likely due to the complex shape and internal heterogeneity of the landslide footprints.~~

- ~~Modigliana: Both U7 and S7 show a tendency to produce False Positives in cultivated fields. These areas are subject to seasonal changes due to agricultural activities such as plowing and vegetation regrowth, which often produce surface variations that resemble the spectral signatures of recent landslides. In particular, disturbed or freshly tilled soil patches in post-event imagery may be incorrectly classified as landslides.~~

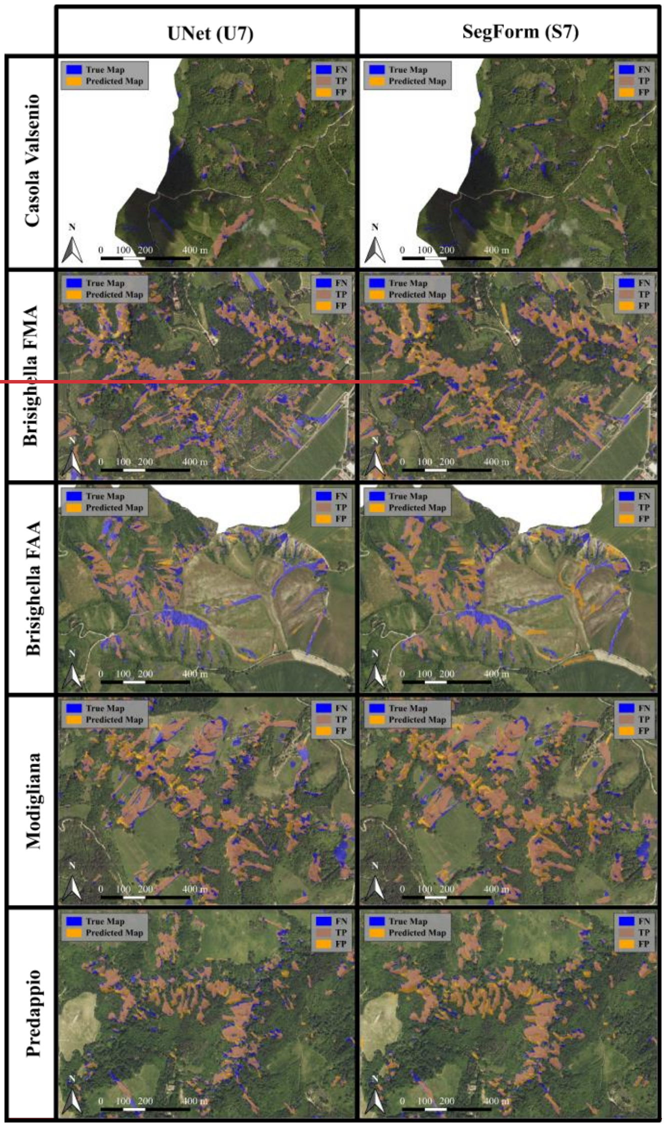
- ~~Brisighella FMA: A different set of issues is observed in this area. False Positives were~~ Figure 10, by contrast, focuses on more challenging scenarios, where the models exhibited recurrent error patterns. In Brisighella FMA, false positives (yellow polygons) are frequently mapped along riverbanks~~, likely caused by localized color differences between the pre- and post-event images. These differences may be~~ and on recently constructed buildings, likely associated with flood-related sediment ~~deposition or changes in water level and channel morphology following localized inundations. Additionally, some False Positives were recorded on recently constructed buildings, especially where bare roofs or fresh concrete surfaces produced~~ redistribution and strong spectral contrasts. ~~These elements, not present in the pre-event imagery, appear as "new" disturbed areas, misleading the models.~~

- ~~Brisighella FAA: This area poses the~~ The most ~~significant challenge for both models. Dominated by badlands~~critical conditions are observed in Brisighella FAA, where widespread false positives occur within badland morphologies developed ~~in~~on Blue Clay formations~~ — geologically and morphologically distinct from the Marnoso-Arenacea Formation of the training area — Brisighella FAA yielded high rates of False Positives. The stark lithological difference appears to have~~, highlighting the limited ~~the models' ability to generalize to this terrain, underlining the importance of including lithological variability in training datasets.~~

- transferability of models trained predominantly on flysch-dominated terrains. In Modigliana, false positives (yellow polygons) are often observed in plowed fields, where agricultural disturbance produces spectral signatures similar to recent landslides. Finally, in Predappio~~: In this area, errors are primarily due to landslides occurring~~, false negatives (blue polygons) dominate in shadowed ~~regions of the post-event imagery. The lower contrast and altered color profiles in these zones~~slopes, where reduced ~~the models' ability to correctly detect landslide features, leading to partial detection or complete omission (False Negatives). These results suggest that terrain shading, common in steep valleys, remains a limitation for optical-image-based mapping.~~

~~Together, these examples illustrate both the strengths and the current limitations of the models when applied in real-world scenarios. While debris-type landslides in well-illuminated, vegetated settings were reliably mapped, generalization to areas with different lithologies, land use, or~~ illumination ~~conditions remains an area for improvement~~and altered colour profiles limit the detectability of landslide features.

Formattato: Rientro: Sinistro: 0 cm, SpazioPrima: 6 pt, Dopo: 6 pt



~~**Figure 8.**~~ By explicitly contrasting successful detections (Figure 9) with failure modes (Figure 10), these visual examples clarify how different surface conditions, illumination effects, and lithological settings influence model performance, supporting the quantitative trends reported in Table 5 and Figure 8.

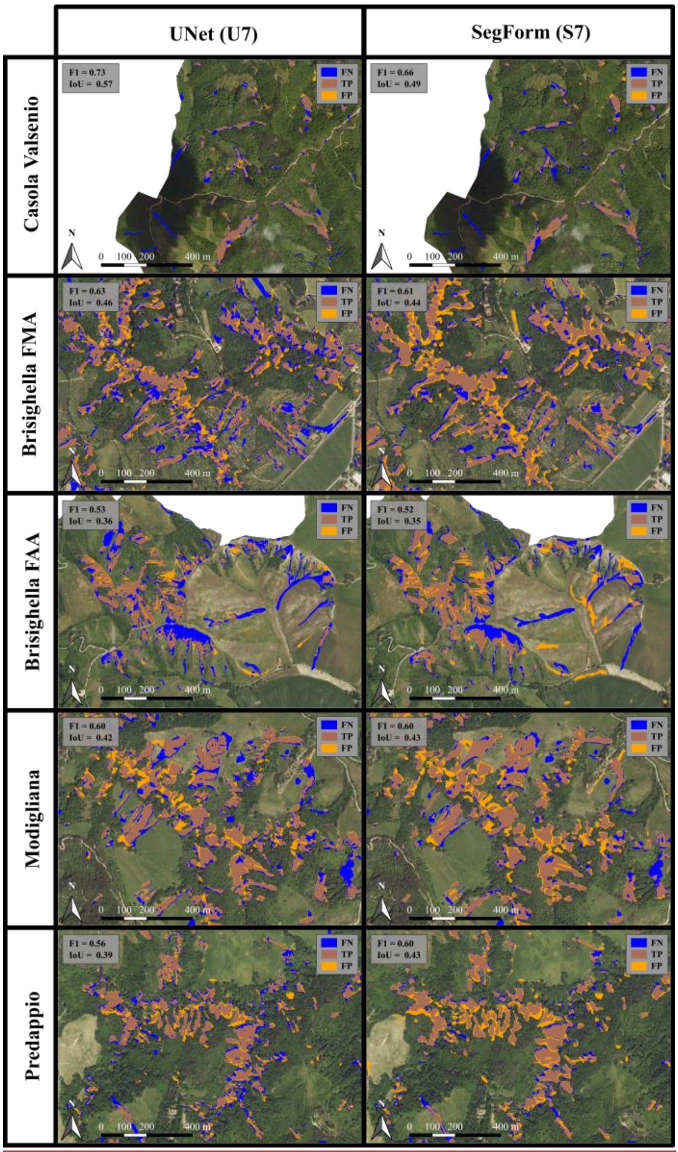
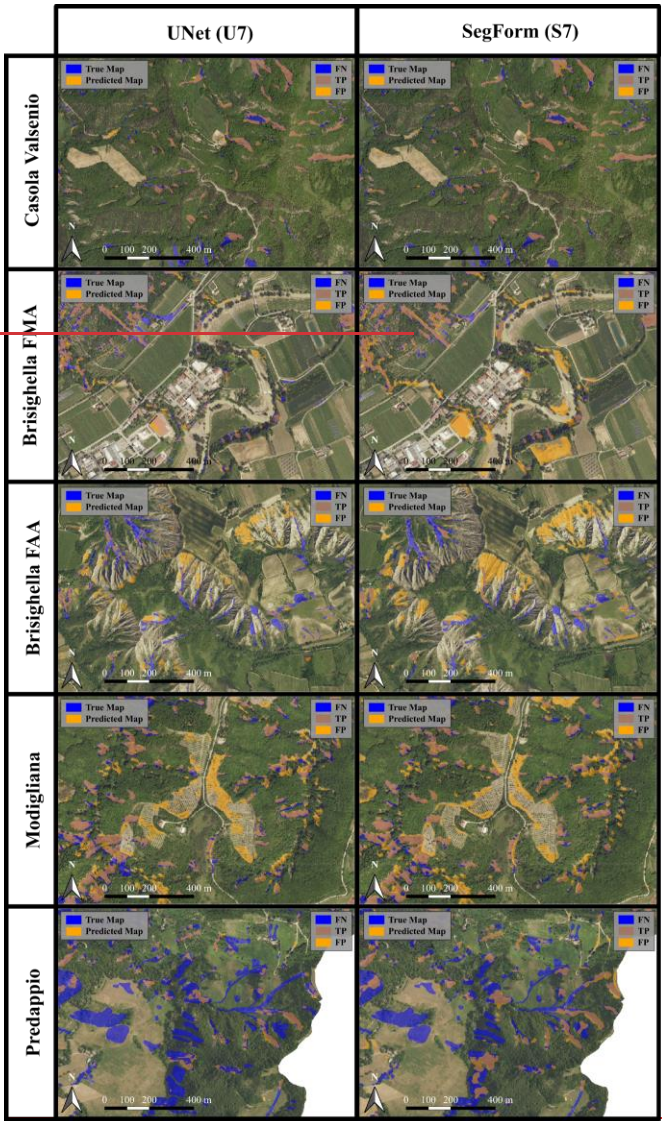


Figure 9. Comparative analysis of landslide mapping results using the U7 (U-Net) and S7 ~~models~~(SegFormer) configurations across ~~different~~the analysed municipalities~~.~~ (Casola Valsenio, Brisighella FMA, Brisighella FAA, Modigliana and Predappio). The figure ~~illustrates~~shows the spatial distribution and extent of ~~landslides as identified by the U7 (UNet-based) and S7 (SegFormer-based) models in Casola Valsenio, Predappio, Modigliana, Brisighella FMA, and Brisighella FAA. Both models were trained using pre- and post-event imagery, slope data, and NDVI from CGR. The comparison highlights the similarities and~~

~~differences in landslide detection capabilities between the two models, particularly in areas dominated by Debris Slides (DS).~~the mapped landslides for both models. In the overlays, the reference inventory (True map) is shown in blue and the model prediction in yellow; their intersection is shown in brown (true positives), while blue-only and yellow-only areas correspond to false negatives and false positives, respectively.



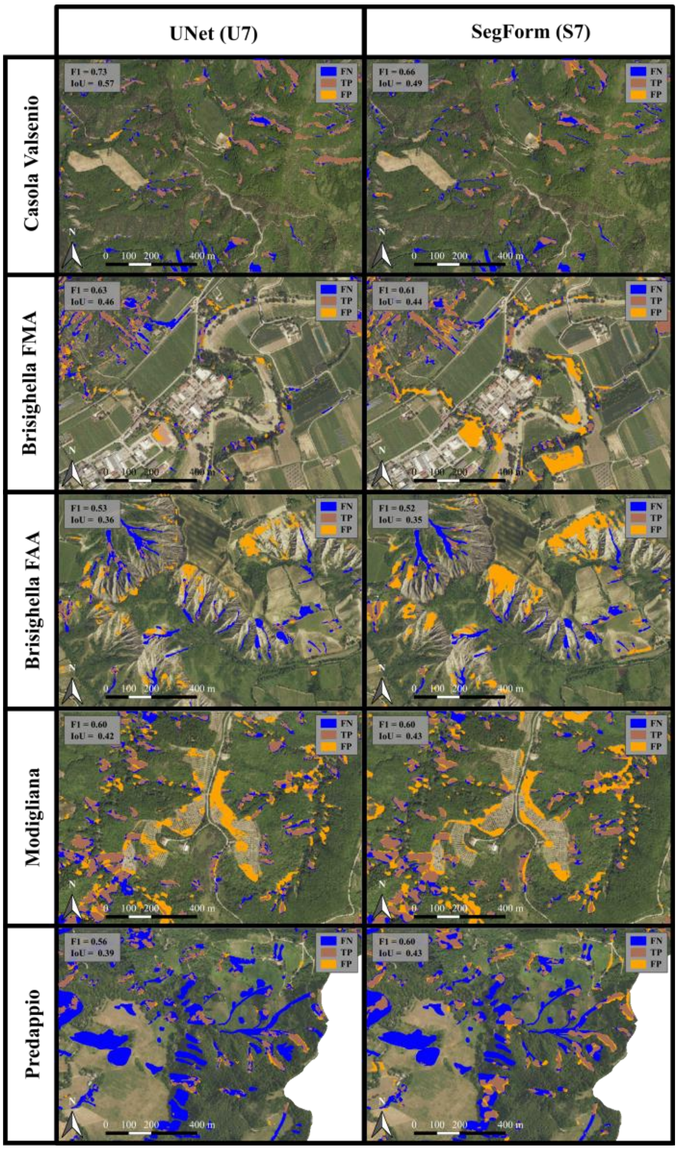


Figure ~~9~~10. Representative examples of challenging scenarios for landslide detection using the U7 (U-Net) and S7 ~~models~~(SegFormer) configurations across municipalities. ~~This figure illustrates common situations where both models encounter difficulties in accurately mapping landslides, highlighting the limitations and potential areas for improvement in the machine learning approach.~~ The ~~examples are drawn from Casola Valsenio, Predappio, Modigliana, Brisighella FMA, and Brisighella FAA, showcasing the diverse geological and~~ panels highlight recurrent error patterns under various environmental conditions ~~that can impact~~

~~model performance~~: Brisighella FMA shows riverbanks and buildings as False Positives (FP), Brisighella FAA displays both False Positives and False Negatives (FP & FN) in badlands, Modigliana FN is observed on plowed fields, and Predappio FN occurs under shadows.

~~4~~5.2 Expert judgement

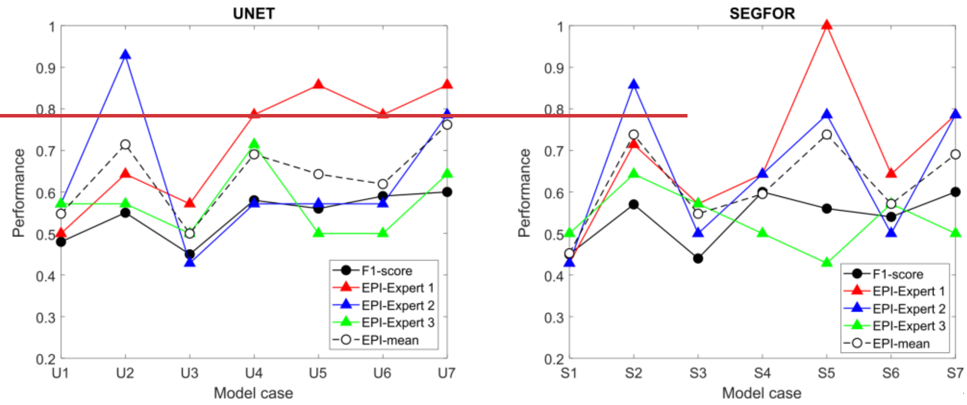
ha formattato: Sottolineato

Figure ~~10 presents a comparison between~~11 compares the quantitative performance (F1-~~scores~~score) of all model ~~cases and~~configurations with the ~~EPI (~~Expert Performance Index~~) values~~ (EPI) derived from ~~expert assessments (see~~ the independent assessment of three experts (Section 3.7). The EPI values ~~shown in the charts~~ reflect ~~evaluations from three experts, who assessed the automated landslide maps produced by the two models and~~ the severity scores assigned ~~weights~~ to seven ~~common detection~~ recurrent mapping errors~~. Table 5 provides the detailed error-specific scores for each case~~ (E1–E7), which were aggregated and normalized to obtain a single index per model configuration.

~~In general, the F1-scores align well with the EPI values obtained from expert judgement. Although the F1-scores exhibit limited variability, the trend of improved performance with increasing model complexity is also shown by the expert-based evaluations. Notably, several performance fluctuations are similarly interpreted: model U3 is clearly recognized as underperforming, while model U7 is consistently considered one of the best by both quantitative scores and expert opinion.~~

~~Some notable discrepancies emerge for certain models that were rated as the best by a single expert but not by F1-scores or the other experts. For instance, model S5 received~~Across configurations, EPI values show the same overall ranking observed for the F1-scores (Fig. 11). While the F1-scores span a relatively narrow interval across cases (Table 5), the EPI values exhibit a comparable monotonic increase from simpler to more information-rich input configurations, indicating that the expert-based assessment captures the same performance gradient. In particular, configuration U3 is associated with one of the lowest EPI values and one of the lowest F1-scores, whereas U7 consistently falls within the top-performing group in both metrics (Fig. 11).

Differences among individual experts are evident for some intermediate configurations. Specifically, the highest EPI score ~~from~~for Expert 1~~, model U2 from~~ is assigned to S5, for Expert 2 to U2, and ~~model U4 from~~for Expert 3 to U4, despite these configurations not being uniformly ranked ~~similarly by others. These differences reflect the specific areas each expert emphasized during evaluation, as well as the inherent subjectivity of expert judgment.~~at the top by the other experts or by the F1-scores. However, ~~the trend of the~~ when expert scores are averaged (mean EPI~~(~~; dashed ~~lines~~line in Fig.~~10), which helps reduce individual biases,~~ 11), the resulting trend closely ~~follows that~~matches the F1-score curve, reducing the influence of individual preferences and providing a more robust overall assessment of ~~the F1-score~~map quality.



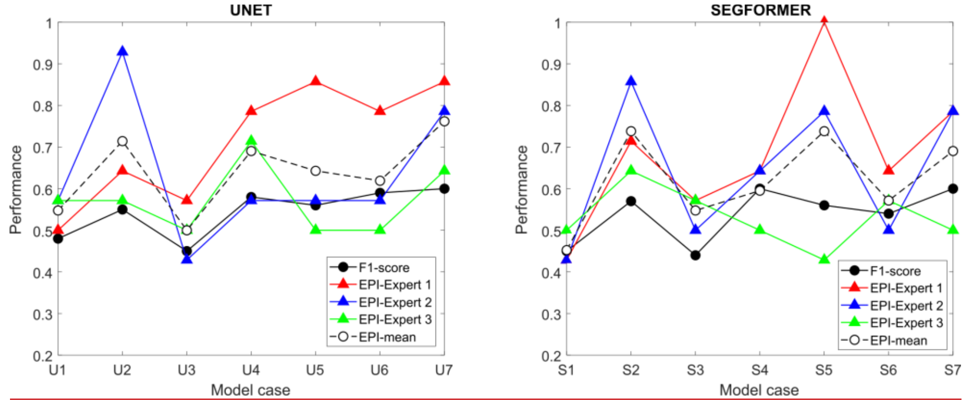


Figure ~~10~~11. Comparison of model performance based on F1-scores and expert-based evaluations from three independent experts.

5.3 Identification of At-Risk Buildings

ha formattato: Tipo di carattere: Non Grassetto, Sottolineato

The identification of at-risk buildings is a crucial component in emergency response scenarios. In the aftermath of the May 2023 disaster, buildings located within landslide boundaries or within a 20-meter buffer zone were considered at risk, influencing subsequent evaluations carried out by the relevant authorities. To assess the effectiveness of our automated landslide mapping for this purpose, we compared the buildings identified by our models to those derived from manual mapping (Pizziolo et al., 2024; Berti et al., 2025).

Figure 12 presents the confusion matrix for the 14 model configurations, showing the comparison between the buildings flagged by our models and those mapped manually. The F1-scores obtained for each case highlight the models' ability to identify structures at risk. For Case 1, U-Net produced an F1-score of 0.53, while SegFormer performed slightly better at 0.67. The highest F1-scores were achieved in Case 7, with U-Net reaching 0.79 and SegFormer 0.75. Notably, models U2, U4, S6, and U7 consistently yielded the best results, with F1-scores averaging around 0.78, demonstrating their effectiveness in accurately identifying buildings at risk.

Despite the relatively high F1-scores, the models still produced both false positives (incorrectly flagged buildings) and false negatives (missed buildings), which underscores the necessity of manual verification in scenarios where precision is critical. However, the performance of the automated models is considerably higher than the manual mapping, especially when considering the large-scale, time-sensitive nature of emergency response. These results suggest that the automated models can significantly assist in the rapid identification of at-risk structures, potentially saving valuable time in the early stages of disaster response. These results are discussed in Section 6.3.

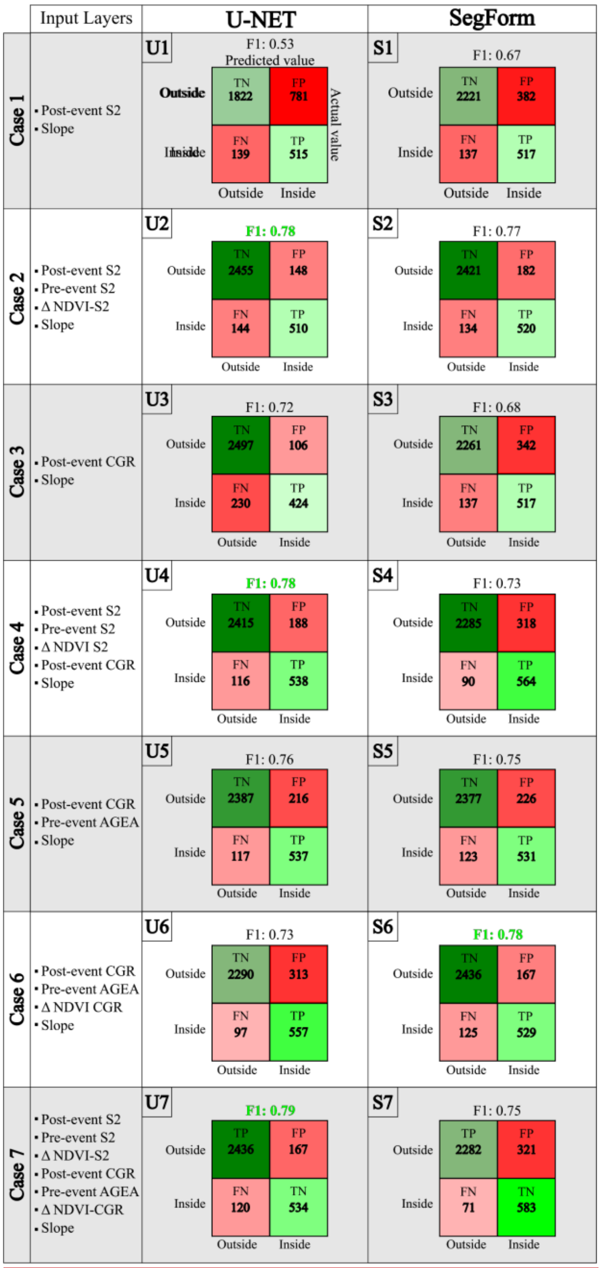


Figure 12. Comparison of buildings identified as at risk (within landslide boundaries or within a 20-meter buffer) by automated mapping methods and manual mapping (ground truth), showing the confusion matrices for all five cases evaluated.

6. Discussion

6.1 Models comparison

One of the key observations from this study concerns the differing sensitivity of U-Net and SegFormer to input data combinations, particularly in the context of emergency response mapping. U-Net's performance varies significantly depending on the layers used, while SegFormer yields more consistent results across different configurations. This contrast is likely rooted in their architectural differences. U-Net, designed for pixel-wise segmentation, relies heavily on the spatial and spectral characteristics of the input data, making it more sensitive to the quality and availability of supplementary layers such as pre-event imagery or NDVI change maps. SegFormer, being a transformer-based model, is better equipped to capture global context, thus reducing dependency on specific input combinations and improving robustness under complex geological conditions.

This architectural contrast becomes even more evident when models are transferred across different geological domains. As shown in the results (Figure ~~7~~8), U-Net exhibits a marked decline in performance when moving from areas geologically similar to the training site (e.g., Predappio, Modigliana) to Brisighella FAA, where lithological conditions differ significantly. This suggests that U-Net has limited generalization capacity beyond the Marnoso-Arenacea Formation (FMA), and struggles in fine-grained, clay-rich terrains like those found in the FAA domain. SegFormer, although generally more stable, also underperforms in Brisighella FAA, highlighting the broader challenge of transferring models across heterogeneous geological settings.

The feature importance values for U-Net and SegFormer (Figure 13) reveal notable differences in how each model prioritizes the input channels. Upon examining the results, we can observe the following key points:

- SegFormer places the highest importance on the post-event CGR channels, particularly the Red and NIR bands, followed by Sentinel-2 Red and Sentinel-2 NIR. This makes sense in the context of post-event analysis, as Red and NIR bands are sensitive to vegetation and surface changes, which are crucial for detecting variations in the landscape after a landslide event. The reliance on these bands is likely due to their ability to capture the shift in vegetation cover (from green to beige or brown) after a disturbance, which can be a key indicator of landslide activity. The CGR (post) channels are particularly effective at highlighting these changes, making them essential for SegFormer's decision-making process.

- U-Net, on the other hand, follows a similar pattern but with lower variability in feature importance across different channels. While the CGR (post) channels remain important, U-Net seems to show a more uniform distribution of feature importance across the input channels, indicating that the model is less sensitive to specific features like Red and NIR. This could be a result of the model's pixel-wise segmentation approach, which might focus more on spatial relationships within the image rather than relying heavily on spectral differences between the bands.

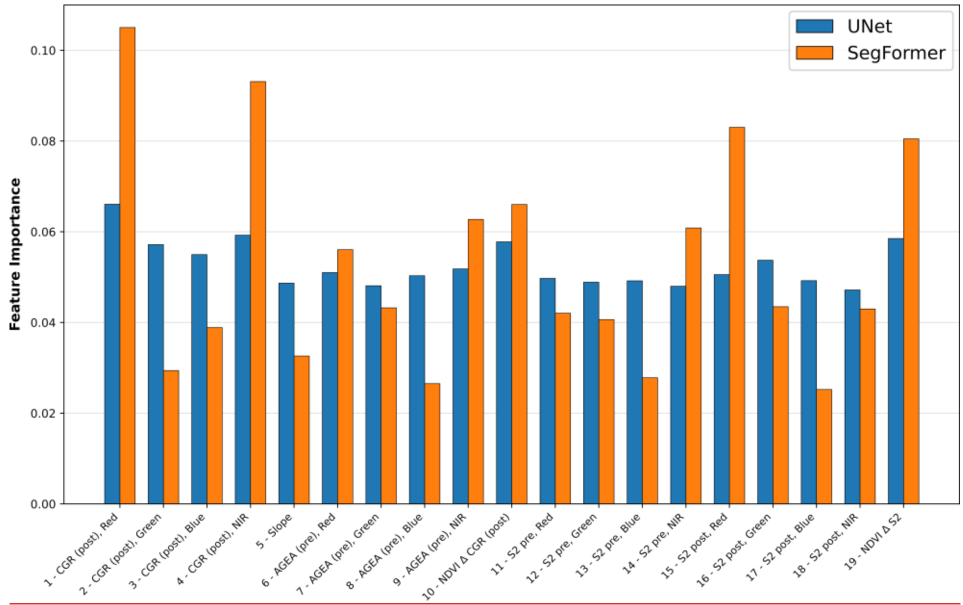


Figure 13. Feature importance values for U-Net and SegFormer across different input channels. The bar chart compares the relative importance assigned to each channel by both models. The blue bars represent U-Net, while the orange bars represent SegFormer. The analysis shows that SegFormer tends to assign higher importance to the post-event CGR channels, particularly the Red and NIR bands, while U-Net exhibits more uniform importance across the channels, with a slightly more localized focus. This difference in feature importance highlights the varying strengths of the models in utilizing spectral information.

Both models exhibit a similar trend in emphasizing the post-event data, particularly the CGR (post) channels. However, SegFormer seems to make use of a wider range of channels, including Sentinel-2 Red and Sentinel-2 NIR, which suggests that SegFormer is more adept at capturing a broader, global context. This capability allows it to effectively integrate multiple data sources. In contrast, U-Net appears to rely more heavily on the specific channels it uses, with its performance being more sensitive to the quality and selection of those channels.

This difference in how the models handle feature importance reflects their underlying architectural differences. SegFormer, with its transformer-based design, is able to capture larger-scale patterns and dependencies within the data, giving it the flexibility to work with diverse and comprehensive input combinations. On the other hand, U-Net's pixel-wise segmentation approach tends to focus on more localized patterns, which can make it less adaptable when handling a variety of input channels.

Nevertheless, both models encountered certain challenges that influenced their performance, which we will now address in the following section.

6.2 Challenges Scenarios

The models in this study encountered several challenging scenarios related to the environmental conditions of the test sites. These challenges, including riverbank areas affected by flooding, geological variations, plowed land, and shadowed regions, directly influenced the models' ability to detect landslides accurately. Spectral analysis of each scenario (Figure 14) reveals distinct patterns that explain the models' performance limitations, while Figure 10 displays the corresponding mapping results with recurrent error patterns.

a) Riverbank Areas Post-Flooding

The May 2023 flood events caused spectral signatures in riverbank areas that partially overlap with those of landslides (Figure 14a). A comparison of pixels from non-landslide riverbank areas and those affected by landslides reveals that while the spectral signatures are similar, there are noticeable differences, particularly in the CGR RGB channels. These differences, although subtle, are appreciable enough to help differentiate the two types of areas. Additionally, the lower slope values in riverbank areas, due to their gentler topography along watercourses, and reduced vegetation change indicators (NDVI Δ channels) set them apart from genuine slope failures. Post-flood sedimentation and bank erosion generate surface disturbances that resemble landslide scars, especially where floodwaters have stripped vegetation and redistributed sediments. Despite these similarities, these features were not sufficiently weighted during training on the Casola Valsenio dataset, leading to false positives in flood-affected environments. However, the distinction in the CGR RGB channels, along with the differences in slope and NIR values, played a crucial role in limiting false negatives (FN) along riverbanks, particularly in the best models (Caso 7, S7, and U7), despite the error still being present.

b) Geological Differences (FMA vs FAA)

The geological difference between Marnoso-Arenacea (FMA) and Blue Clays (FAA) proved challenging. Pixels from landslides in both lithologies were compared, with notable differences in the CGR (Red) and CGR (NIR) channels, which are critical for landslide detection (Figure 14b). FAA regions exhibit lighter colors in both pre- and post-event imagery, indicating typical erosion and sediment redistribution in badlands (see Figure 10). This aligns with the diluvial processes observed in the Blue Clays region. Training the models solely on FMA lithology, such as Casola Valsenio, rendered them ineffective when applied to FAA regions, where the spectral characteristics differ significantly. This highlights the limited transferability of models trained on a single lithology.

c) Plowed Fields (Modigliana)

Spectral signatures from landslide-affected areas in Modigliana were compared with those from plowed fields outside the landslide zones (Figure 14c). The similarities in spectral signatures, particularly in the CGR (Blue) and CGR (Green) channels, led to false positives (yellow polygons) in Figure 10. Agricultural fields, when plowed, share similar characteristics with disturbed landslides, especially when considering the differences in pre-event imagery. These differences are more pronounced in the Slope and pre-event channels. However, due to the relatively low importance assigned to these channels during training on the Casola Valsenio dataset, the models struggled to differentiate these fields from landslides, resulting in confusion and false positives in plowed fields.

d) Shadow Effects (Predappio)

Shadowed regions, particularly in Predappio, made it nearly impossible for the models to classify these areas as landslides. The spectral analysis (Figure 14d) highlights much darker values in shadowed areas in the CGR (post-event) imagery, which complicates the recognition of landslides. This is due to the drastically reduced reflectance in shadowed pixels, which prevents the models from detecting landslides in these regions. Fortunately, this issue is isolated to a small portion of Predappio. The reduced illumination and altered color profiles in shadowed areas hinder the model’s ability to detect the land movement, causing false negatives in these regions, as shown in Figure 10.

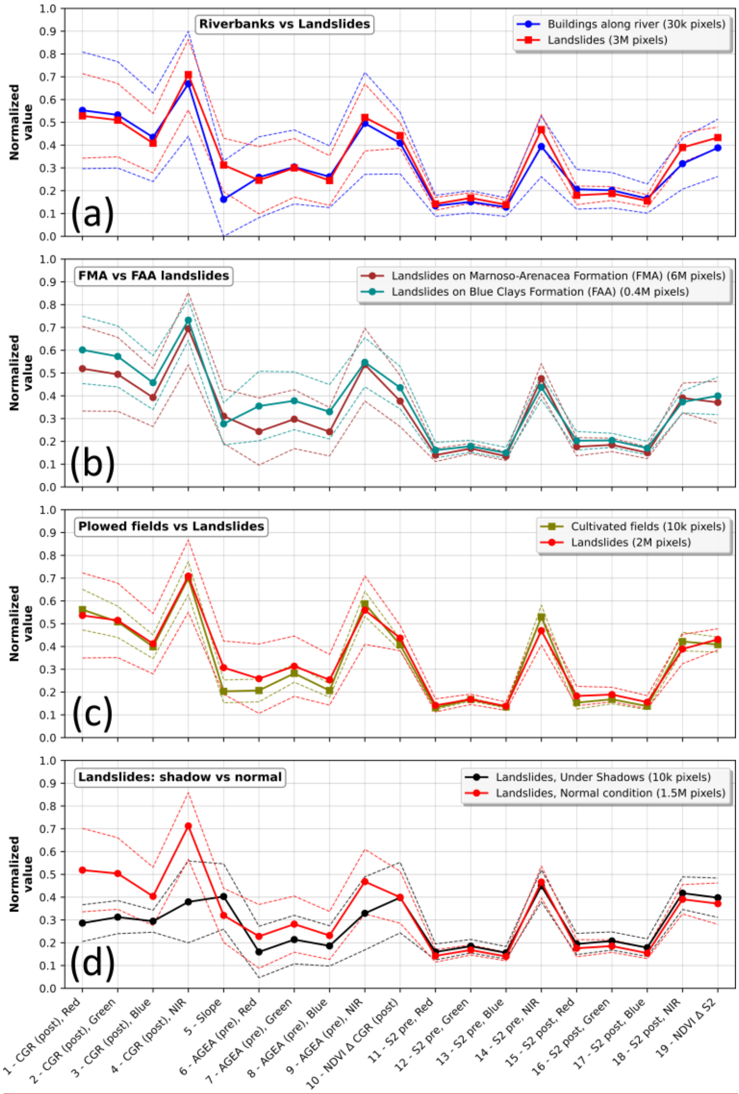


Figure 14. Spectral analysis of different land types and landslide detection. (**a**) Comparison between riverbank areas affected by flooding and landslides. The blue line represents the spectral characteristics of buildings along the river, while the red line corresponds to landslides. (**b**) Spectral differences between landslides occurring on Marnoso-Arenacea Formation (FMA) and Blue Clays (FAA), showing the significant spectral dissimilarity. (**c**) Comparison between spectral signatures of plowed fields and landslides. The green line represents plowed fields, while the red line corresponds to landslides, with noticeable overlap in spectral characteristics. (**d**) Impact of shadowed regions on

landslide detection, comparing normal conditions (red) with shadowed regions (black). The graph reveals how shadows significantly reduce the ability of the models to detect landslides.

The analysis of various challenging scenarios, including flood-affected riverbanks, geological variations, plowed land, and shadowed regions, reveals a consistent pattern: while identifiable spectral discriminants exist for each challenge, the models' training on a single geological and environmental domain (Casola Valsenio, FMA lithology) limited their ability to generalize across varying conditions. The failure to adapt to different geological contexts, land use patterns, flood-affected areas, and illumination conditions highlights the importance of domain-adaptive approaches and more geographically comprehensive training datasets for robust landslide detection across diverse landscapes.

To further investigate the role of geology in model generalization, an additional analysis was conducted to assess the impact of including lithology among the input layers. For ~~instance, comparing~~this, the lithological layer was categorized into 8 distinct classes, with each class corresponding to a unique Unit ID (as detailed in Table 2). These classes were numerically encoded from 1 to 8, representing different lithological units. We compared the performance of Case S3 (RGB + slope) with the same configuration plus ~~a~~the lithology layer. However, this did not yield significant improvements in Casola, Predappio, Modigliana, or Brisighella MA~~. These~~, as these areas are predominantly characterized by the Marnoso-Arenacea Formation ~~(“~~("Unit 7~~”),~~"), which ~~is~~was already well represented in the training dataset (Figure 2a~~), meaning~~). As a result, the addition of the lithology layer ~~provides~~provided redundant information. However, when the same model was applied to Brisighella FAA, where the lithology is dominated by Blue Clays ~~(“~~("Unit 1~~”), it~~"), the model failed to detect any landslides, resulting in a very low F1-score of 0.02. This ~~highlights~~underscores a key limitation: simply adding lithology as a static input layer is insufficient for ensuring generalization. If the training dataset is not lithologically balanced, the model may reinforce existing biases, associating landslide occurrence primarily with ~~“~~"Unit 7~~”~~" and failing to detect events in ~~“~~"Unit 1~~”.~~".

This finding ~~underscores~~highlights a fundamental ~~point~~issue in landslide detection models: effectively incorporating lithological information ~~into AI-based landslide mapping~~ requires more than ~~providing~~just adding it as an input feature. The training dataset itself must include landslide polygons from a representative range of lithological settings. When the model was explicitly trained and tested on Brisighella FAA using the same configuration (Case S3: RGB + slope), the F1-score increased from 0.32 to 0.41, confirming that geological consistency in the training data ~~substantially~~significantly improves the ~~model's~~model's ability to detect landslides in previously underrepresented domains.

Although the F1-scores and Intersection over Union (IoU) achieved in this study may appear modest compared to ~~some~~generic machine~~-~~-learning benchmarks ~~(Papers with Code), they represent a substantial accomplishment in~~ (e.g., Ghorbanzadeh et al., 2021, 2022; Prakash et al., 2021; Meena et al., 2023), it is important to ground these results within the context of landslide detection and segmentation literature, where lower scores are often observed. Several factors contribute to these relatively low F1-scores, including class imbalance between landslide and non-landslide areas, the spatial heterogeneity of landslides, uncertainties in reference inventories, and the influence of image resolution and labeling quality. Landslide mapping~~. Unlike~~, unlike conventional image classification tasks, ~~landslide mapping~~ requires ~~the~~ precise delineation of ~~complex and~~ irregular and complex shapes across highly variable terrain~~. Our models aim to produce a realistic representation of landslide extents, including subtle transitions and intricate morphologies, rather than simplified or generalized outputs. This level of~~, which makes it more challenging to achieve high scores.

Our dataset was specifically designed for emergency response scenarios, where a small portion of the affected area (Casola Valsenio) is mapped and the model is applied to the rest. This led to a train-validation-test ratio of 12% for training, 3% for validation, and 85% for testing. The limited training and validation sets, combined with the complex nature of landslide mapping, contribute to modest F1-scores. Such a trade-off between accuracy ~~is~~ and generalization is common in remote sensing, where high-resolution, fine-grained feature extraction is both critical and difficult to ~~attain, yet essential for practical applications in emergency response and hazard assessment~~achieve at a large scale.

~~In this regard~~Nevertheless, the potential of these automated approaches in emergency scenarios ~~could be~~is considerable. ~~The ability to~~ By rapidly ~~generate~~generating landslide maps of comparable quality to expert-drawn products, these methods could ~~greatly~~significantly accelerate initial response efforts. ~~By reducing~~Reducing the need for time-consuming manual digitization~~, these methods~~ could save weeks of work, ~~an especially critical advantage~~which is crucial during crisis events (Berti et al~~.~~., 2025). The automatically generated maps may serve as a robust initial product, ~~allowing~~enabling practitioners to focus on refinement and validation rather than starting from scratch, ~~thereby enabling the delivery of a~~ ultimately delivering high-quality final ~~map~~maps in a fraction of the time required for full manual mapping.

6.3 Identification of At-Risk Buildings

A further consideration in our study is the use of automated landslide maps ~~for identifying~~to identify damage and at-risk structures. Following the May 2023 disaster, the Emergency Commission determined that all buildings located within landslide boundaries or within a 20-~~metre~~meter buffer were ~~to be~~ considered at risk and potentially subject to relocation. To evaluate the ~~usefulness~~effectiveness of our automated mapping products for this task, we compared the buildings identified automatically with those derived from manual mapping.~~ Figure 11 presents the confusion matrix for the 16 modelled cases.~~

The F1-scores for the models indicate that ~~“~~"U2~~”, “~~", "U4~~”, “~~", "S6~~”,~~", and ~~“~~"U7~~”~~" are the most accurate CNN outputs,~~ successfully~~ identifying on average 528 out of 654 buildings at risk (~0.78 F1-score). These results demonstrate the ~~models’~~models' ability to estimate the spatial extent of the phenomenon, even when using Sentinel-2 imagery at ~~10 m~~10 m resolution (U2). ~~Nevertheless~~However, all models produced both false positives (incorrectly flagged buildings) and false negatives (~~overlooked~~missed buildings~~). Consequently,~~), underscoring the need for manual verification ~~remains essential~~ when high accuracy is required~~, and misclassifications cannot be tolerated~~. Even a small number of misclassified structures, especially inhabited buildings, can have serious consequences in emergency situations, thus highlighting the importance of improving model performance for more effective response efforts.

~~Even a small number of misclassified structures may have serious consequences when inhabited buildings are involved. This raises a key question: how can model performance be further improved~~

~~to support more effective and precise emergency response efforts in the future? Several avenues could be explored:~~

1. ~~Higher resolution input data: As discussed in Section 3.1, high-resolution aerial imagery (20 cm/pixel) and NDVI change maps were down sampled to 2 m/pixel to reduce computational load. Using the original resolution, while computationally expensive, could yield improved detail and model performance (Wang et al., 2022).~~
2. ~~Improved tiling strategies: The current approach relies on fixed 496×496 px tiles for training, validation, and testing (Section 3.5). This can split landslide polygons at tile boundaries and reduce spatial context. As noted by Abrahams et al. (2024), employing sliding windows or full-scene processing could mitigate these issues and enhance segmentation quality.~~
3. ~~Data augmentation: No augmentation techniques were used in this study. Simple operations such as flipping, rotation, or brightness adjustment, shown by Prakash et al. (2021) could improve CNN robustness reducing overfitting and improve generalizability.~~

~~Finally, although current approaches enable rapid mapping of landslides, their ability to assess structural exposure remains limited. Buildings located just outside mapped polygons may still be at significant risk, which a uniform 20 m buffer may fail to capture adequately. A more advanced solution is proposed by BGC Engineering (2010), who incorporate topographic features and estimated landslide runout directions to dynamically assess exposure. In this framework, risk is evaluated based on terrain morphology, flow paths, and downslope connectivity, offering a more realistic representation of potential impacts compared to distance-based buffers. Such models could represent a major step forward in automated damage assessment, complementing segmentation outputs with geomorphologically informed prioritization strategies.~~

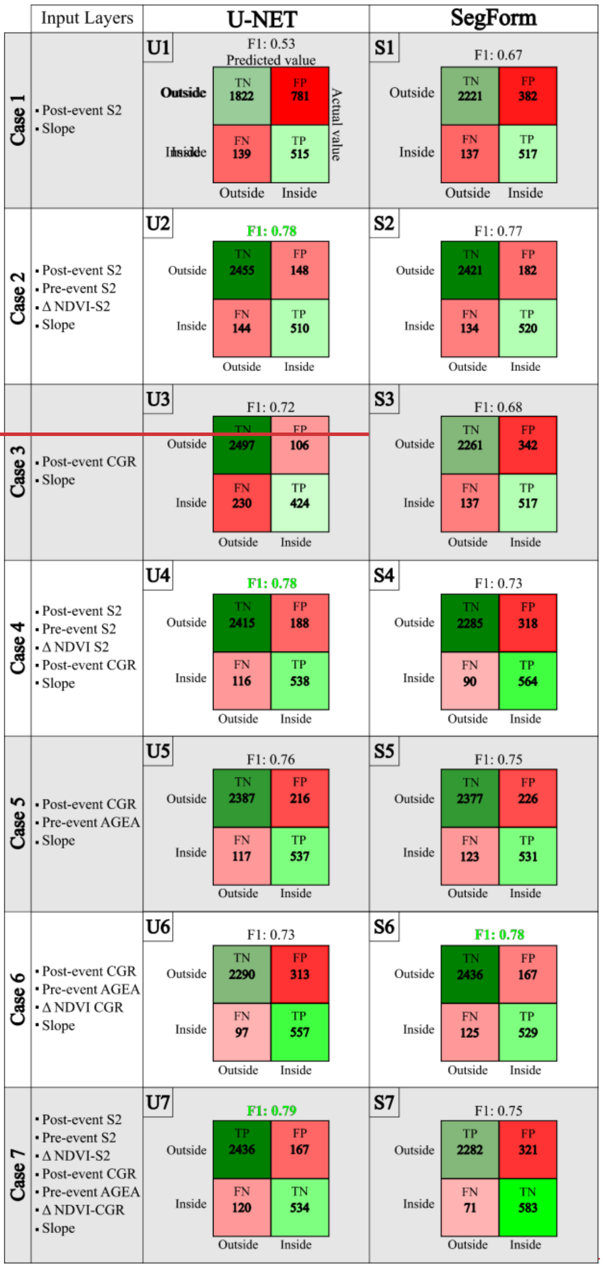


Figure 11. ~~Comparison of buildings identified as at risk (within landslide boundaries or within a 20-meter buffer) by automated mapping methods and manual mapping (ground truth), showing the confusion matrices for all five cases evaluated.~~

~~6.~~Improving the landslide mapping itself is crucial for improving the accuracy of at-risk building identification. As discussed in Section 7, refining mapping methodologies will lead to more reliable and consistent results in future disaster scenarios.

6.4 Future Research Directions

To overcome the limitations identified in this study, future research should focus on key areas. A primary challenge is optimizing data acquisition to improve map quality, including minimizing shadow effects by scheduling imagery collection around noon, especially during winter months. Expanding training datasets to include more plowed field examples will help models better differentiate agricultural lands from landslides, reducing false positives. Further refinement could include detailed land-use layers or slope filters to improve accuracy.

Improving data resolution is another important direction. While this study used 2-meter resolution data, higher-resolution datasets (e.g., 20 cm) would provide more detailed information, enabling better detection. Leveraging higher computational power would be necessary to process these datasets and enhance model performance. In terms of approach, future research could explore advanced techniques such as Multiscale Feature Pyramid Networks (FPN), which allow models to process multiple scales simultaneously, Graph Neural Networks (GNNs) for capturing complex spatial relationships in geospatial contexts, and Neural Architecture Search (NAS) to automatically identify the best network architecture, improving robustness and generalization (Lin et al., 2017; Kipf & Welling, 2017; Zoph & Le, 2017).

Lastly, addressing the issue of geologically diverse training datasets remains crucial. Training models on datasets with limited geological diversity can hinder generalization. Expanding datasets to include various lithologies and geomorphological features will enhance model robustness. As deep learning models evolve, it is essential to prioritize the collection of diverse, high-quality data, as even the most advanced models cannot replace the need for comprehensive datasets. Future research should focus on improving data acquisition, enhancing model generalization, and refining validation techniques to further strengthen AI-based mapping in disaster management.

7. Conclusion

Formattato

~~In this work, we~~This study investigated the potential of automated landslide mapping to support rapid emergency response following the extreme meteorological events of May 2023 in Emilia-Romagna, Italy. Using a deep learning approach, we trained and tested two semantic segmentation models, U-Net and SegFormer, on high-resolution aerial imagery, Sentinel-2 ~~imagery~~data, NDVI change maps, and slope data. The training was conducted solely in the municipality of Casola Valsenio, while model performance was assessed on three additional municipalities (~~Predappio~~Brisighella, Modigliana, ~~Brisighella~~Predappio), chosen for their geological settings and significant landslide occurrence.

~~The main findings of the study can be summarized as follows:~~

- ~~**Training performance**: ~~In the training area of Casola Valsenio, both ~~models ~~U-Net and SegFormer achieved high performance~~ when provided~~, with ~~rich ~~consistent results across different input combinations. ~~U-Net reached F1 scores around 0.70, particularly when both~~ The models showed no significant variation in performance based on the input layers, as long as high-resolution post-event imagery ~~and NDVI change maps were ~~(CGR) was included ~~(Figure 7). These results confirm that, when trained on spatially detailed and spectrally rich datasets, CNN-based models can closely replicate manually mapped landslide boundaries, including small and morphologically complex features.~~

- ~~**Generalization to test areas**: When ~~. However, when applied to ~~unseen areas, both U-Net and SegFormer produced comparable F1 scores across Predappio, ~~external regions such as Brisighella, Modigliana, and ~~Brisighella MA. While slight variations were observed, SegFormer occasionally performing better in certain configurations, there was no systematic advantage indicating stronger generalization. In Brisighella FAA, both models experienced a clear performance drop, highlighting the challenge of transferring models to geologically distinct terrains (Table 4, Figure 7). These results suggest that generalization is primarily constrained by lithological dissimilarity, rather than by architectural differences between CNN and transformer-based models~~Predappio, the models experienced a decrease in generalization. Despite this, the automated mapping still successfully identified the majority of landslides, demonstrating its utility in emergency scenarios.

Formattato: Nessun elenco puntato o numerato

- ~~**Model comparison**: Despite their architectural differences, the two models achieved similar performance across all test ~~In areas~~. This convergence likely reflects the limited variability and discriminative power of the available input layers, suggesting that both architectures were able to extract the most relevant features from the data (Figure 7). Improving the variety and informativeness of the input layers could unlock further improvements, especially in geologically heterogeneous regions.~~

- ~~**F1 vs. Expert Index**: The Expert Performance Index (EPI) was largely consistent with the quantitative F1 scores, reinforcing the validity of standard segmentation metrics for assessing model quality (Figure 10). Experts noted that the best-performing models also produced outputs that were more coherent from a geomorphological perspective, but they also identified recurring issues such as over segmentation, misclassification in agricultural areas, and false positives along riverbanks. These errors, while systematic, are easily corrected during manual review, making the model highly valuable as a starting point for rapid emergency mapping.~~

- ~~**Lithology influence**: The inclusion of lithology as an additional input layer had limited effect when the model was applied to areas already represented in the training dataset. However, in~~ like Brisighella FAA~~,~~ (Blue Clays), where relevant lithologies were ~~almost absent from~~underrepresented in the training set, ~~both~~the models ~~performed poorly unless explicitly retrained~~struggled to generalize effectively. This highlights ~~that effective integration of lithological information into AI-based mapping requires more than adding it as an input feature: the training dataset must include landslide polygons across a representative~~the importance of incorporating a diverse range of ~~lithological contexts. When the model was retrained on Brisighella FAA, the F1 score increased, confirming that geological consistency in the training data significantly improves ~~lithologies in the training data to ensure robust model generalization ~~performance~~across different geologically complex regions.

Formattato: Nessun elenco puntato o numerato

- ~~**Exposure and damage detection**: The automated maps successfully identified a majority of the buildings located within or near landslide polygons, reaching an F1 score of approximately 0.78 for building-at-risk detection (Figure 11). However, both false positives (e.g., buildings wrongly marked as at risk) and false negatives (e.g., undetected at-risk buildings) were frequent. This limits the reliability of automated exposure assessments, particularly in densely populated or infrastructure-rich areas. More advanced exposure analysis methods that account for slope direction, runout potential, and terrain morphology (e.g., BGC Engineering, 2010) are needed to complement segmentation-based approaches.~~

~~Overall, automated landslide mapping offers substantial advantages in emergency scenarios. While not a substitute for manual efforts, it provides a reliable baseline that can significantly accelerate the production of actionable maps. With further refinement, such as full-resolution imagery (Wang et al., 2022), improved tiling strategies (Abrahams et al., 2024), and data augmentation (Prakash et al., 2021), these tools can become critical assets in disaster management workflows. Given the increasing scale and frequency of climate-driven disasters, integrating AI-based approaches into civil protection protocols could enhance both the speed and consistency of emergency responses. Deep learning models, despite current limitations, represent a scalable solution to the growing demand for rapid, accurate, and spatially detailed hazard information. By automatically addressing the most evident features, these models reduce the burden on experts, allowing them to focus on verifying complex or ambiguous regions. Nonetheless, manual validation remains essential, as AI outputs alone cannot yet ensure sufficient reliability in sensitive contexts.~~

~~A key recommendation in such extreme events is the prompt acquisition of high-resolution imagery immediately after the triggering event. As demonstrated by the Emilia-Romagna Region's response to the May 2023 floods, early access to detailed aerial data improved mapping accuracy and supported timely disaster response. In that case, the Italian government submitted a request to the European Union Solidarity Fund (European Commission, 2024), leading to a proposed allocation of over €97 million. In future emergencies, timely and accurate damage assessments, potentially supported by automated tools, could play a strategic role in accelerating disaster documentation and accessing financial recovery mechanisms.~~

The choice of model architecture, whether U-Net or SegFormer, had minimal impact on performance, with both models showing similar results. Performance decreased with less detailed inputs, like Sentinel imagery alone, and improved with more rich inputs, reinforcing that data quality, not architecture, drives performance.

In emergency contexts, where external data for validation may not be available, expert judgment becomes essential for selecting the most reliable maps when ground truth data is lacking. While automated maps offer rapid assessments, they often require expert adjustments to refine outputs offering a balanced approach between speed and accuracy. The correct and thoughtful integration of AI-based systems into civil protection protocols represents a critical step forward, ensuring that these technologies complement existing procedures and improve overall response effectiveness.

Formattato

Acknowledgements

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan ~~–~~- NRRP, Mission 4, Component 2, Investment 1.3 ~~–~~- D.D. 1243 2/8/2022, PE0000005)

Disclosures and declarations

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4 (chat.openai.com) to enhance the grammar and syntax, as well as to refine the sentence structure. All the content is original, and no concepts, ideas, or interpretations were produced by this tool. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Formattato: whitespace-pre-wrap, SpazioDopo: 8 pt, Interlinea: multipla 1,15 ri

~~Funding~~Code and~~/or Conflicts~~ data availability

ha formattato: Tipo di carattere: Grassetto, Nessuna sottolineatura

The dataset used in this study is openly available at https://doi.org/10.5281/zenodo.13742643.
The code developed for the analyses is available from the authors upon reasonable request.

Author contribution

NDS performed the data analysis, developed the scripts, and wrote the manuscript.
GC contributed to the interpretation of ~~interests/~~the results and supervised the work.

ha formattato: Nessuna sottolineatura

DE contributed to the methodology and algorithmic framework.
ELP contributed to the methodology and algorithmic framework.
AC supervised the interpretation of the results.
MB conceived the study, supervised the analysis and manuscript preparation, and provided critical review and guidance throughout the research process.

Competing interests

ha formattato: Tipo di carattere: Grassetto, Nessuna sottolineatura

The authors declare they have no financial interests.

References

Abrahams, E., Snow, T., Siegfried, M. R., & Pérez, F. (2024). A Concise Tiling Strategy for Preserving Spatial Context in Earth Observation Imagery. *arXiv preprint arXiv:2404.10927*.

~~Alaska Satellite Facility (ASF) Distributed Active Archive Center. (2023). *ALOS PALSAR data*. NASA DAAC. Retrieved from https://search.asf.alaska.edu~~

Amatya, P., Scheip, C., Déprez, A., et al. (2023). Learnings from rapid response efforts to remotely detect landslides triggered by the August 2021 Nippes earthquake and Tropical Storm Grace in Haiti. Nat Hazards, 118, 2337–2375. https://doi.org/10.1007/s11069-023-06096-6

Amy, L. A., & Talling, P. J. (2006). Anatomy of turbidites and linked debrites based on long distance (120 × 30 km) bed correlation, Marnoso Arenacea Formation, Northern Apennines, Italy. *Sedimentology, 53*(1), 161–212. https://doi.org/10.1111/j.1365-3091.2005.00756.x

Ban, Y., Zhang, P., Nascetti, A. *et al.* Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Sci Rep* **10**, 1322 (2020). https://doi.org/10.1038/s41598-019-56967-x

Berti, M., Pizziolo, M., Scaroni, M~~., Generali, M., Olivucci, S., Gozza, G., et al. (2024). Emergency mapping: lessons learned from 2023 landslide event in Romagna. Manuscript submitted for publication in Landslides.~~. et al. Testing NDVI and U-Net for automated mapping of multiple-occurrence regional landslide events using satellite and aerial multispectral data (Casola Valsenio, Emilia-Romagna, Northern Apennines, Italy). *Landslides* (2026). https://doi.org/10.1007/s10346-025-02671-z

Berti, M., Pizziolo, M., Scaroni, M., Generali, M., Critelli, V., Mulas, M., Tondo, M., Lelli, F., Fabbiani, C., Ronchetti, F., Ciccarese, G., Dal Seno, N., Ioriatti, E., Rani, R., Zuccarini, A., Simonelli, T., & Corsini, A. (2025). RER2023: The landslide inventory dataset of the May 2023 Emilia-Romagna meteorological event. *Earth System Science Data, 17*(3), 1055–1074. https://doi.org/10.5194/essd-17-1055-2025

Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 65(1), 2–1. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Brath, A., Casagli, N., Marani, M., Mercogliano, P., & Motta, R. (2023). Report by the Technical-Scientific Commission established by Regional Council Resolution No. 984/2023 and Managerial Determination No. 14641/2023 to analyze the extreme meteorological events of May 2023. Technical-Scientific Commission Report. Retrieved June 15, 2024, from https://www.regione.emilia-romagna.it/alluvione/

Casagli, N., Cigna, F., Bianchini, S., et al. (2016). Landslide mapping and monitoring by using radar and optical remote sensing: examples from the EC-FP7 project SAFER. Remote Sens Appl Soc Environ, 4, 92–108.

Catani, F., Tofani, V., & Raspini, F. (2020). Use of remote sensing for landslide studies. Progress in Landslide Science, 75-102.

Chen, Z., Zhang, Y., Ouyang, C., Zhang, F., & Ma, J. (2018). Automated landslides detection for mountain cities using multi-temporal remote sensing imagery. Sensors (Basel), 18, 821. https://doi.org/10.3390/s18030821

ha formattato: Italiano (Italia)

ha formattato: Inglese (Stati Uniti)

ha formattato: Inglese (Stati Uniti)

ha formattato: Inglese (Stati Uniti)

ha formattato: Inglese (Stati Uniti)

Codice campo modificato

Chen, X., Liu, M., Li, D., Jia, J., Yang, A., Zheng, W., & Yin, L. (2023). Conv-trans dual network for landslide detection of multi-channel optical remote sensing images. Frontiers in Earth Science, 11, 1182145. https://doi.org/10.3389/feart.2023.1182145

Cruden, D.M., Varnes, D.J.: Landslide Types and Processes, Transportation Research Board, U.S. National Academy of Sciences, Special Report, 247, 36-75, 1996

Dal Seno, N., Evangelista, D., Piccolomini, E., Berti, M. Comparative analysis of conventional and machine learning techniques for rainfall threshold evaluation under complex geological conditions. Landslides (2024). https://doi.org/10.1007/s10346-024-02336-3

~~District of North Vancouver. (2009). *Landslide risk assessment for select escarpment slopes*. BGC Engineering Inc. https://www.geoweb.dnv.org/Products/Reports/PublicSafety/BGCLandslideRiskAssessment_Final.pdf~~

Dosovitskiy, A., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

~~European Commission. (2024). *Proposal for a Decision of the European Parliament and of the Council on the mobilisation of the European Union Solidarity Fund to provide assistance to Italy*. COM/2024/325 final. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=COM:2024:325:FIN~~

~~European Space Agency (ESA). (2019). *Copernicus Digital Elevation Model (DEM)*. https://doi.org/10.5270/ESA-c5d3d65~~

~~Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... & Alsdorf, D. E. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2).~~

Ferrario, M. F., & Livio, F. (2023). Rapid Mapping of Landslides Induced by Heavy Rainfall in the Emilia-Romagna (Italy) Region in May 2023. Remote Sens (Basel), 16, 122. https://doi.org/10.3390/rs16010122

Foraci, R., Tesini, M. S., Nanni, S., Antolini, G., & Pavan, V. (2023). L’inquadramento meteo e idrologico degli eventi. Ecoscienza, ARPAE Emilia-Romagna, anno XIV, 5, 20-24.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning Convolutional Neural Networks for landslide detection. Remote Sensing, 11(2), 196. ~~https://doi.org/10.3390/rs11020196~~https://doi.org/10.3390/rs11020196

Ghorbanzadeh, O., Xu, Y., Ghamisi, P., Kopp, M., & Kreil, D. (2022). Landslide4Sense: Reference Benchmark Data and Deep Learning Models for Landslide Detection. IEEE Transactions on Geoscience and Remote Sensing, 60. https://doi.org/10.48550/arXiv.2206.00515

Ghorbanzadeh, O., Shahabi, H., Crivellari, A., Homayouni, S., Blaschke, T., & Ghamisi, P. (2022). Landslide detection using deep learning and object-based image analysis. Landslides, 19(4), 929-939. https://doi.org/10.1007/s10346-021-01843-x

ha formattato: Legature: Nessuna

ha formattato: Inglese (Stati Uniti)

ha formattato: Inglese (Stati Uniti)

Ghorbanzadeh O., Khalil Gholamnia & Pedram Ghamisi (2023) The application of ResU-net and OBIA for landslide detection from multi-temporal Sentinel-2 images, Big Earth Data, 7:4, 961-985, DOI: 10.1080/20964471.2022.2031544.

Guzzetti, F., Mondini, A. C., Cardinali, M., et al. (2012). Landslide inventory maps: New tools for an old problem. Earth Sci Rev, 112, 42–-66. https://doi.org/10.1016/j.earscirev.2012.02.001.

Holbling, D., Eisank, C., Albrecht, F., et al. (2017). Comparing manual and semi-automated landslide mapping based on optical satellite images from different sensors. Geosciences. https://doi.org/10.3390/geosciences7020037.

Isensee, F., Jaeger, P.F., Kohl, S.A.A. *et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203–-211 (2021). https://doi.org/10.1038/s41592-020-01008-z

Iverson, R. M., George, D. L., Allstadt, K., et al. (2015). Landslide mobility and hazards: implications of the 2014 Oso disaster. Earth Planet Sci Lett, 412, 197–-208. https://doi.org/10.1016/j.epsl.2014.12.020

LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–-444 (2015). https://doi.org/10.1038/nature14539

Lin, T.-Y., Dollár, P., Girshick, R., He, K., & Hariharan, B. (2017). Feature Pyramid Networks for Object Detection. *CVPR 2017*. https://arxiv.org/abs/1612.03144

Liu, X., Zhao, C., Zhang, Q., Lu, Z., Li, Z., Yang, C., Zhu, W., Liu-Zeng, J., Chen, L., & Liu, C. (2021). Integration of Sentinel-1 and ALOS/PALSAR-2 SAR datasets for mapping active landslides along the Jinsha River corridor, China. *Engineering Geology, 284*, 106033. https://doi.org/10.1016/j.enggeo.2021.106033

Lu, W., Hu, Y., Zhang, Z., Chen, L., & Liu, C. (2023). A dual-encoder U-Net for landslide detection using Sentinel-2 and DEM data. *Landslides, 20*(8), 1975–-1987. https://doi.org/10.1007/s10346-023-02089-5

Ji, S. P., Yu, D. W., Shen, C. Y., Li, W. L., & Xu, Q. (2020). Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides, 17, 1337–-1352. https://doi.org/10.1007/s10346-020-01353-2.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. arXiv preprint. https://doi.org/10.48550/arXiv.1412.6980

~~NASA Shuttle Radar Topography Mission (SRTM)(2013). Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF. Accessed: 2025-01-16~~

~~NASA Earthdata Search. (2023). *Earthdata Search Tool*. NASA. Retrieved from https://search.earthdata.nasa.gov/~~

Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. *ICLR 2017*. https://arxiv.org/abs/1609.02907

Nava, L.; Bhuyan, K.; Meena, S.R.; Monserrat, O.; Catani, F. Rapid Mapping of Landslides on SAR Data by Attention U-Net. Remote Sens. 2022, 14, 1449. https://doi.org/10.3390/rs14061449

Notti, D., Cignetti, M., Godone, D., & Giordan, D. (2023). Semi-automatic mapping of shallow landslides using free Sentinel-2 images and Google Earth Engine. *Natural Hazards and Earth System Sciences, 23*(9), 2625~~–~~-2648. https://doi.org/10.5194/nhess-23-2625-2023

Novellino, A., Pennington, C., Leeming, K. *et al.* Mapping landslides from space: A review. *Landslides* **21**, 1041~~–~~-1052 (2024). https://doi.org/10.1007/s10346-024-02215-x

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., & Kumar, K. V. (2010). Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*, 116, 24–36. https://doi.org/10.1016/j.geomorph.2009.10.004

Meena, S. R., Ghorbanzadeh, O., & van Westen, C. J., et al. (2021). Rapid mapping of landslides in the Western Ghats (India) triggered by 2018 extreme monsoon rainfall using a deep learning approach. Landslides, 18, 1937-1950. https://doi.org/10.1007/s10346-020-01602-4

Meena, S. R., Soares, L. P., & Grohmann, C. H., et al. (2022). Landslide detection in the Himalayas using machine learning algorithms and U-Net. Landslides, 19, 1209-1229. https://doi.org/10.1007/s10346-022-01861-3

Meena, S. R., Nava, L., Bhuyan, K., Puliero, S., Soares, L. P., Dias, H. C., Floris, M., and Catani, F. (2023). HR-GLDD: a globally distributed dataset using generalized deep learning (DL) for rapid landslide mapping on high-resolution (HR) satellite imagery, Earth Syst. Sci. Data, 15, 3283~~–~~-3298, https://doi.org/10.5194/essd-15-3283-2023

Mondini, A. C., Santangelo, M., Rocchetti, M., Rossetto, E., Manconi, A., & Monserrat, O. (2019). Sentinel-1 SAR Amplitude Imagery for Rapid Landslide Detection. *Remote Sensing*, *11*(7), 760. https://doi.org/10.3390/rs11070760

Prakash, N., Manconi, A., & ~~Loetangw~~Loew, S. ~~(2021).~~ (2020). Mapping Landslides on EO Data: Performance of Deep Learning Models vs. Traditional Machine Learning Models. Remote Sensing, 12(3), 346. https://doi.org/10.3390/rs12030346

Prakash, N., Manconi, A. & Loew, S. A new strategy to map landslides with a generalized convolutional neural network. Sci Rep~~,~~ 11~~(1), 1-15.~~, 9722 (2021). https://doi.org/10.1038/s41598-021-89015-8

N. Prakash and A. Manconi (2021). Rapid Mapping of Landslides Triggered by the Storm Alex, October 2020. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 1808-1811. https://doi.org/10.1109/IGARSS47720.2021.9553321

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems (pp. 8026-8037).

ha formattato: Tipo di carattere: +Corpo (Calibri), 11 pt, Legature: Standard + Contestuali

ha formattato: Tipo di carattere: Times New Roman, 12 pt, Legature: Nessuna

ha formattato: Tipo di carattere: +Corpo (Calibri), 11 pt, Legature: Standard + Contestuali

ha formattato: Tipo di carattere: Times New Roman, 12 pt, Italiano (Italia), Legature: Nessuna

ha formattato: Tipo di carattere: Times New Roman, 12 pt, Italiano (Italia), Legature: Nessuna

Pizziolo, M., Generali, M., & Scaroni, M. (2023). In Appennino un numero di frane mai riscontrato prima. Ecoscienza, ARPAE Emilia-Romagna, anno XIV, 5, 31-33.

Pizziolo, M., Berti, M., Scaroni, M., Generali, M., Critelli, V., Mulas, M., Tondo, M., Lelli, F., Fabbiani, C., Ronchetti, F., Ciccarese, G., Dal Seno, N., Ioriatti, E., Rani, R., Zuccarini, A., Simonelli, T., & Corsini, A. (2024). RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event - Version 1 (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13742643

Prakash, N., Manconi, A. & Loew, S. A new strategy to map landslides with a generalized convolutional neural network. *Sci Rep* **11**, 9722 (2021). https://doi.org/10.1038/s41598-021-89015-8

Prechelt, L. (1998). Early stopping-but when? In Neural Networks: Tricks of the Trade (pp. 55-69). Springer, Berlin, Heidelberg.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 658-666).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, 234-241.

~~Rosenqvist, A., Shimada, M., Ito, N., & Watanabe, M. (2007). ALOS PALSAR: A Pathfinder Mission for Global-Scale Monitoring of the Environment. *IEEE Transactions on Geoscience and Remote Sensing, 45*(11), 3307–3316. https://doi.org/10.1109/TGRS.2007.901027~~

Sameen, M. I., & Pradhan, B. (2019). Landslide detection using residual networks and the fusion of spectral and topographic information. IEEE Access, 7, 114363–114373. https://doi.org/10.1109/access.2019.2935761

Tang, X., Tu, Z., Wang, Y., Liu, M., Li, D., & Fan, X. (2022). Automatic Detection of Coseismic Landslides Using a New Transformer Method. Remote Sens, 14, 2884. https://doi.org/10.3390/rs1412f2884

Tharwat, A. (2020). Classification assessment methods. Applied Computing and Informatics, 17(1), 168-192.

~~Wang, Y., Bashir, S. M. A., Khan, M., Ullah, Q., Wang, R., Song, Y., ... & Niu, Y. (2022). Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Systems with Applications*, *197*, 116793.~~

Wasowski, J., & Bovenga, F. (2014). Investigating landslides and unstable slopes with satellite Multi Temporal Interferometry: Current issues and future perspectives. *Engineering Geology, 174*, 103–138. https://doi.org/10.1016/j.enggeo.2014.03.003

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). Transformers: State-of-the-art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38-45).

Wu, L., Liu, R., Ju, N., Zhang, A., Gou, J., He, G., & Lei, Y. (2024). Landslide mapping based on a hybrid CNN-transformer network and deep transfer learning using remote sensing images with topographic and spectral features. International Journal of Applied Earth Observation and Geoinformation, 126, 103612. https://doi.org/10.1016/j.jag.2023.103612

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Advances in Neural Information Processing Systems, 34, 12077-12090.

Xu, Y., Ouyang, C., Xu, Q. *et al.* CAS Landslide Dataset: A Large-Scale and Multisensor Dataset for Deep Learning-Based Landslide Detection. *Sci Data* **11**, 12 (2024). https://doi.org/10.1038/s41597-023-02847-z

Yang, W., Wang, Y., Sun, S. *et al.* Using Sentinel-2 time series to detect slope movement before the Jinsha River landslide. *Landslides* **16**, 1313–1324 (2019). https://doi.org/10.1007/s10346-019-01178-8

Ye CM et al (2019) Landslide detection of hyperspectral remote sensing data based on deep learning with constrains. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12:50475060. https://doi.org/10.1109/JSTARS.2019.2951725.

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net, in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749-753, May 2018. doi:10.1109/LGRS.2018.2802944.

Zoph, B., & Le, Q. V. (2017). Neural Architecture Search with Reinforcement Learning. *ICLR 2017*. https://arxiv.org/abs/1611.01578

ha formattato: Italiano (Italia)

ha formattato: Italiano (Italia)