ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE DELLA TERRA, DELLA VITA E DELL'AMBIENTE

Ciclo 37

Settore Concorsuale: 04/A3 - GEOLOGIA APPLICATA, GEOGRAFIA FISICA E GEOMORFOLOGIA

Settore Scientifico Disciplinare: GEO/05 - GEOLOGIA APPLICATA

APPLICATION OF MACHINE LEARNING METHODS FOR LANDSLIDE RISK MITIGATION

Presentata da: Nicola Dal Seno

Coordinatore Dottorato

Barbara Cavalazzi

Supervisore

Matteo Berti

Co-supervisore

Elena Loli Piccolomini

Esame finale anno 2025

Abstract

Landslides in complex geological settings are a widespread and dangerous natural hazard that, in the context of climate change, poses significant challenges for modern society. Despite knowledge about triggering mechanisms and monitoring techniques, conducting comprehensive studies of landslide susceptibility and emergency management in regions like Emilia-Romagna, Italy, often entails significant obstacles. This PhD project analyzes the multifaceted aspects of landslide prediction and management in this geologically complex region.

The research initially focused on improving rainfall thresholds for forecasting, comparing empirical-statistical approaches with machine learning (ML) techniques such as XGBoost, Random Forest, and neural networks. Results demonstrated that ML models outperformed traditional methods, achieving higher predictive accuracy and reducing false positives. While these techniques proved effective, challenges such as data quality and interpretability were highlighted, particularly in operational contexts. A key innovation was the development of a unified platform integrating ML-based rainfall thresholds into operational early warning systems, providing real-time risk assessments.

The catastrophic rainfall events of May 2023 in Emilia-Romagna posed a significant challenge to our research, necessitating an unprecedented rapid response. These occurrences triggered thousands of landslides, underscoring the importance of swift and precise mapping for effective emergency management. Considering this urgent need, a multi-institutional collaboration was quickly established, leading to a comprehensive landslide inventory documenting 80,000 polygons using high-resolution aerial imagery. While invaluable for immediate recovery planning and future risk assessment, the time-consuming nature of manual mapping underscored the pressing need for faster, automated solutions. Our research pivoted to explore rapid automated mapping techniques, starting with trials using the U-Net neural network in severely affected municipalities like Casola Valsenio. More advanced algorithms, such as SegFormer, were subsequently employed across diverse settings including Modigliana, Predappio, and Brisighella. These methods demonstrated their ability to rapidly process large datasets, achieving high levels of accuracy in identifying landslide-affected areas. This urgent application of our research offered valuable insights into the practical challenges of rapid landslide response, informing innovations for future methodologies.

In conclusion, this research demonstrates the transformative potential of integrating machine learning with traditional approaches in landslide prediction and management. Practical lessons learned include the importance of high-quality data, balancing interpretability with accuracy, and combining automated methods with expert validation for effective deployment in emergency contexts. The improved prediction models, early warning platform, comprehensive

landslide inventory, and rapid mapping techniques collectively contribute to building resilient communities capable of effectively responding to and mitigating the impacts of landslides in the face of increasing climate-related risks.

CONTENTS

Chapter 1

Introduction

1.1 Preface

This chapter provides an overview of the primary themes and methodologies addressed in this PhD research. The study investigates the application of Machine Learning (ML) methods for landslide risk mitigation and explores how ML can improve disaster response efficiency, particularly during emergency situations.

Firstly, an introduction to the fundamental concepts of machine learning will be presented, including key definitions, algorithms, and learning approaches (section 1.21). This will be followed by an overview of the current state-of-the-art technologies in the field, with particular attention to deep learning, convolutional neural networks (CNNs), and their practical applications in civil protection (section 1.3).

Section 1.3 delves into the real-world applications of ML for landslide prediction and mapping. Special focus will be given to the integration of ML algorithms in predicting rainfall-induced landslides and automated mapping during emergencies, which are essential for early warning systems and damage assessment (section 1.3.1).

Finally, section 1.4 presents the research questions that guided the development of this thesis and outlines the structure of the following chapters, providing a roadmap for the reader through the subsequent analysis and findings.

1.2 Machine Learning: A Comprehensive Introduction

1.2.1 Defining Machine Learning

Machine Learning (ML) has become a revolutionary force in computer science, reshaping how we approach problem-solving and data-driven decision-making. Defined by Arthur Samuel in 1959 as "the field of study that gives computers the ability to learn without being explicitly programmed," ML has evolved dramatically. It allows systems to learn and improve from vast amounts of data, bypassing the need for manually coded instructions. This paradigm shift, particularly the introduction of neural networks, has fundamentally changed our ability to tackle complex problems. As Jordan (1996) observed, neural networks have unlocked new capacities for machines to process data in ways previously unimaginable, drawing on past experiences to improve over time.

ML's ability to adapt to new data is especially crucial in areas where traditional rule-based programming struggles. For instance, in fields like spam detection, where manually crafted rules can quickly become outdated, ML models can continuously adjust their strategies by learning from new data, leading to more effective results (Bhowmick & Hazarika, 2018). This adaptability has shown significant benefits in domains that rely on large datasets, such as medicine and astrophysics, where ML is used to predict patient outcomes or analyze vast amounts of astronomical data, identifying patterns that would otherwise go unnoticed (Ball & Brunner, 2010). These breakthroughs highlight ML's capacity to process high-dimensional data and extract valuable insights, making it a critical tool in a range of scientific and commercial applications.

Despite the increasing success of ML in industries such as finance, e-commerce, and healthcare, its application in fields like civil protection, territorial management, and disaster response has been slower. This measured adoption reflects the complexity of integrating ML into real-world emergency management scenarios. While ML has the potential to improve how we manage and respond to natural disasters, particularly by processing real-time data and providing more accurate forecasts, it is important to recognize that these improvements may be context-specific. In some cases, ML models can assist in predicting events like landslides or floods, optimizing resource allocation, and enhancing early warning systems. These contributions could play a role in minimizing the impact of such events, but they are not a universal solution (Sun et al., 2019). The successful application of ML in these areas also depends on overcoming significant challenges, particularly related to the quality and availability of data. In disaster scenarios, data is often incomplete or biased, which can limit the performance of ML systems. Ensuring that datasets are diverse, representative, and up-to-date is crucial to making informed decisions and avoiding the risk of exacerbating inequalities during emergencies (Gebru et al., 2018).

In contrast, the private sector has quickly embraced ML, using it to streamline operations, foster innovation, and enhance customer experiences. Industries like finance and technology rely on

ML to analyze large datasets and make data-driven decisions, which has given these sectors a significant competitive edge (Brynjolfsson & McAfee, 2017). However, the public sector faces distinct challenges in implementing ML, including budgetary constraints, regulatory frameworks, and the necessity to ensure equitable service delivery. Unlike private companies that can focus on profitability, public institutions must prioritize fairness, transparency, and accountability in the use of ML (Sun et al., 2019; Mikhaylov et al., 2018).

Nonetheless, ML holds great potential for improving disaster management. By analyzing real-time data, ML could improve the accuracy of forecasts and optimize resource deployment during emergencies. To fully realize this potential, public sector organizations must address barriers related to data quality, computational resources, and ethical considerations. Overcoming these challenges will be essential for integrating ML into emergency response systems, helping to build more resilient societies that are better equipped to anticipate and mitigate the impacts of natural disasters.

In conclusion, ML represents a fundamental shift from explicit programming to adaptable, data-driven learning. Its ability to analyze complex datasets and uncover valuable insights has already transformed industries worldwide. However, its untapped potential in fields like civil protection and disaster management presents a significant opportunity. With the right frameworks and careful attention to data quality and ethical use, ML could become an essential tool for improving how societies respond to and mitigate the impacts of natural disasters, enhancing both preparedness and resilience.

1.2.2 Machine Learning Timeline

Machine Learning (ML) has evolved through decades of research, innovation, and practical application. This journey, spanning over seven decades, has been characterized by conceptual breakthroughs, technological leaps, and paradigm shifts. This timeline not only charts the technical milestones but also reflects the changing relationship between human cognition and machine capability, offering crucial insights into the future trajectory of AI (Figure 1.1).

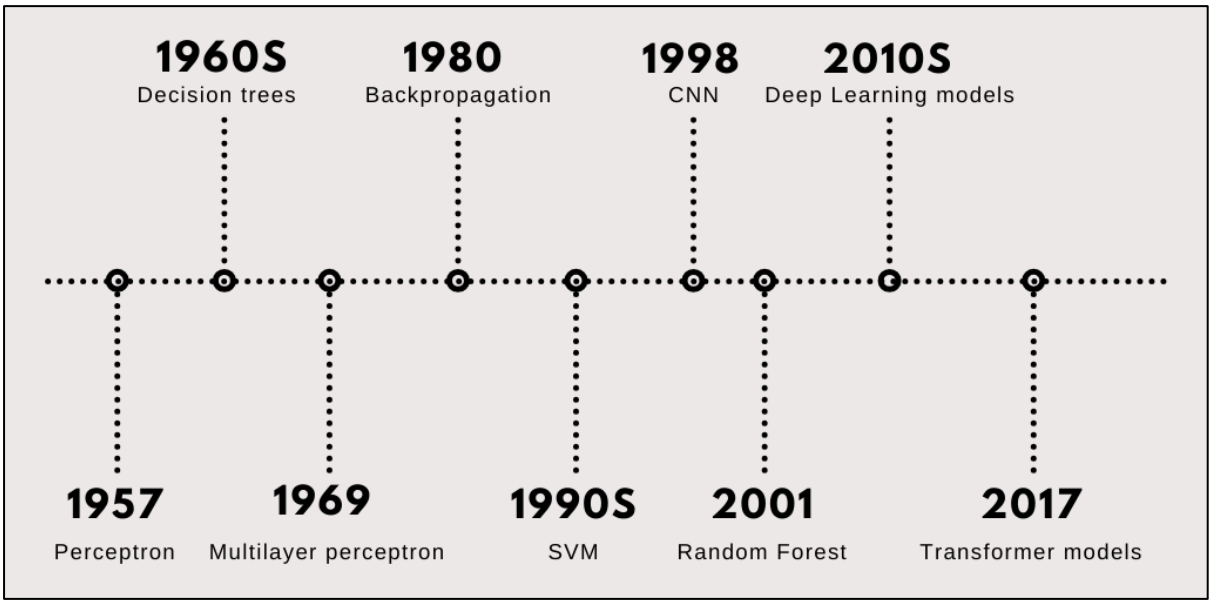


Figure 1.1. Important Machine Learning milestones timeline (modified from Medium).

1.2.2.1 Early Foundations (1950s-1960s)

The concept of machine learning began to take shape in the 1950s, a period that also marked the dawn of artificial intelligence. During this time, foundational ideas were introduced that would profoundly influence the development of the field. In 1950, the now-famous Turing Test was proposed as a method for assessing a machine's ability to exhibit intelligent behavior indistinguishable from that of a human (Turing, 1950). This idea laid the groundwork for exploring how machines could emulate human thought processes.

A few years later, a groundbreaking checkers-playing program was developed that could improve its performance through self-play, representing an early demonstration of what would later be known as reinforcement learning (Samuel, 1959). This work was one of the first instances of machine learning from experience, an essential concept in the field today.

The late 1950s also saw significant advancements in both neural network theory and statistical learning. In 1957, the perceptron was introduced (Figure 1.1), a simple type of neural network that became the foundation for more complex deep learning models developed decades later (Rosenblatt, 1958). That same year, a fundamental method for clustering data known as K-means was proposed, though it wasn't formally published until 1982 (Lloyd, 1982). Additionally, in 1958, Logistic Regression was introduced, further expanding the range of statistical tools available for machine learning applications (Cox, 1958). These early contributions set the stage for the rapid evolution of machine learning, laying the theoretical and practical foundations that continue to underpin the field.

1.2.2.2 A.I. Winter (1970s-1980s)

The initial enthusiasm for artificial intelligence was followed by a period of diminished interest and reduced funding, often referred to as the "AI Winter." Despite this challenging environment, significant theoretical advancements occurred during this time. In 1975, the concept of backpropagation was introduced (Figure 1.1), which became an essential algorithm for training neural networks (Werbos, 1974). The 1980s also saw the emergence of decision trees and other statistical learning methods, with pioneering work in decision tree algorithms having a lasting impact on the field (Quinlan, 1986). These contributions laid the groundwork for future breakthroughs in AI, despite the broader stagnation experienced during this period.

1.2.2.3 Advancements in Machine Learning Algorithms (1990s-2010s)

The evolution of machine learning in the 1990s marked a significant shift as researchers began to embrace statistical methods and introduce innovative algorithms. Among the key breakthroughs of this period was the development of the Random Forests algorithm in 1995, which revolutionized classification and regression tasks by utilizing an ensemble of decision trees (Ho, 1995). Around the same time, the introduction of Support Vector Machines (SVMs) by (Figure 1.1) provided a powerful new framework for tackling classification problems, offering a robust method that could handle both linear and non-linear data (Cortes & Vapnik, 1995).

As the field advanced into the late 1990s and early 2000s, there was a notable resurgence of interest in neural networks, largely fueled by significant improvements in existing techniques. The backpropagation algorithm, which had been conceptualized in the 1970s, was refined and became central to the effective training of neural networks, thanks in part to the influential work of LeCun et al. (1998). This period also saw the introduction of boosting techniques, such as AdaBoost, which addressed key limitations of earlier models by improving their accuracy and reducing bias (Freund & Schapire, 1997).

The late 1990s also witnessed the development of Long Short-Term Memory (LSTM) networks in 1997, a breakthrough that overcame the vanishing gradient problem that had hindered the effectiveness of earlier recurrent neural networks (Hochreiter & Schmidhuber, 1997). In 1998, the potential of convolutional neural networks (CNNs) was demonstrated (Figure 1.1), who showcased their efficacy in image recognition tasks, setting the stage for the deep learning revolution that was to come (LeCun et al., 1998). This momentum continued to build into the 2000s, introduction of deep belief networks in 2006, marking a pivotal moment that reignited widespread interest in deep learning and solidified its role in modern AI (Hinton et al., 2006).

1.2.2.4 The Deep Learning Revolution (2010s-Present)

The last decade has witnessed unprecedented advancements in ML, particularly in deep learning. A pivotal moment came in 2012 when AlexNet, won the ImageNet competition, marking a turning point in computer vision (Krizhevsky et al., 2012). Two years later, the introduction of Generative Adversarial Networks (GANs), opened new possibilities in generative modeling (Goodfellow et al., 2014).

The field of natural language processing was revolutionized in 2017 with the introduction of the transformer architecture (Vaswani et al., 2017). This paved the way for large language models like BERT and GPT, which demonstrated remarkable capabilities in various language tasks between 2018 and 2020 (Devlin et al., 2019; Brown et al., 2020).

Recent years have seen the mainstream adoption of large language models, exemplified by ChatGPT in 2022. The focus has also intensified on few-shot and zero-shot learning, with models demonstrating improved performance with minimal task-specific data (Brown et al., 2020). Concurrently, the importance of Explainable AI has grown, with new techniques being developed to interpret complex deep learning models (Samek et al., 2023).

The narrative of Machine Learning is far from a linear progression; it is a complex interplay of theoretical advances, technological innovations, and societal needs. Each breakthrough has expanded the boundaries of what machines can learn and accomplish, often in ways unforeseen by their creators. As ML continues to evolve, it not only pushes the frontiers of artificial intelligence but also challenges our fundamental conceptions of intelligence, creativity, and the human-machine relationship. The future of ML promises not just technological advancement, but a profound reimagining of our cognitive landscape.

1.2.3 Learning Approaches

Machine Learning (ML) encompasses a diverse set of techniques that enable computers to learn from data and make informed decisions. These techniques can be broadly categorized into several key approaches: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and transfer learning (Figure 1.2). Each approach has its unique characteristics, strengths, and ideal use cases, depending on the nature of the data and the problem at hand. Understanding these different learning paradigms is essential for selecting the appropriate method for a given task and for developing robust ML models that can effectively address a wide range of real-world challenges. The following sections provide an overview of these fundamental learning approaches, highlighting their core principles and applications.

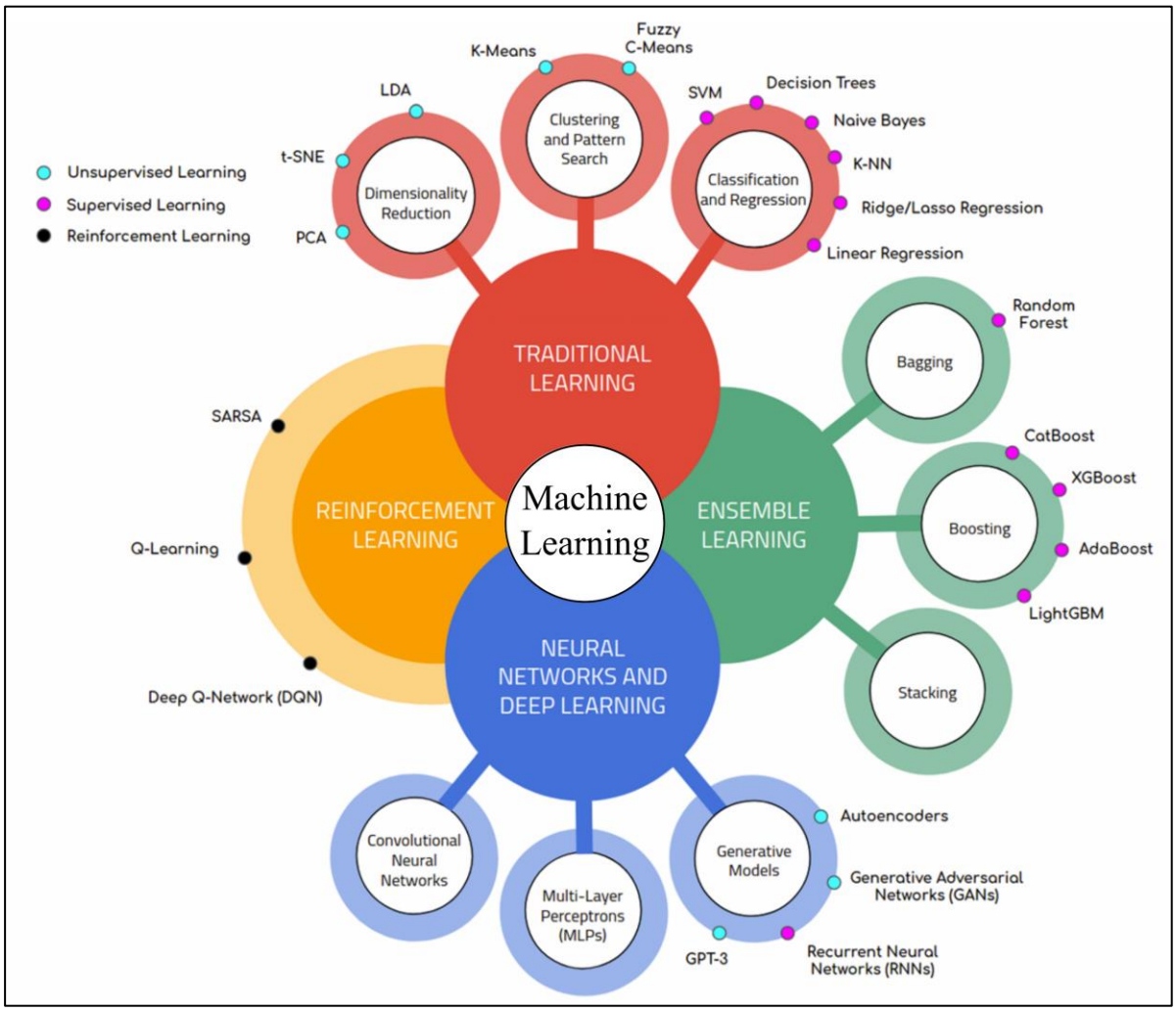


Figure 1.2. Conceptual scheme of various machine learning approaches and relative algorithms. (Karthikeyan et al., 2022, modified).

1.2.3.1 Supervised Learning

Supervised learning is perhaps the most common and well-understood approach in ML. In this paradigm, the algorithm learns from labeled data, where both input features and corresponding target variables are provided. The goal is to learn a function that maps inputs to correct outputs, enabling the model to make predictions on new, unseen data (Figure 1.3). Supervised learning is further divided into classification tasks, where the output is a discrete category (such as spam detection or image classification), and regression tasks, where the output is a continuous value (like price prediction or age estimation) (Murphy, 2012). Common algorithms in this category include linear regression, logistic regression, decision trees, support vector machines, and many neural network architectures (Hastie et al., 2009). The strength of supervised learning lies in its ability to make accurate predictions when provided with high-quality labeled data, making it invaluable in applications ranging from medical diagnosis to financial forecasting (Bishop, 2006).

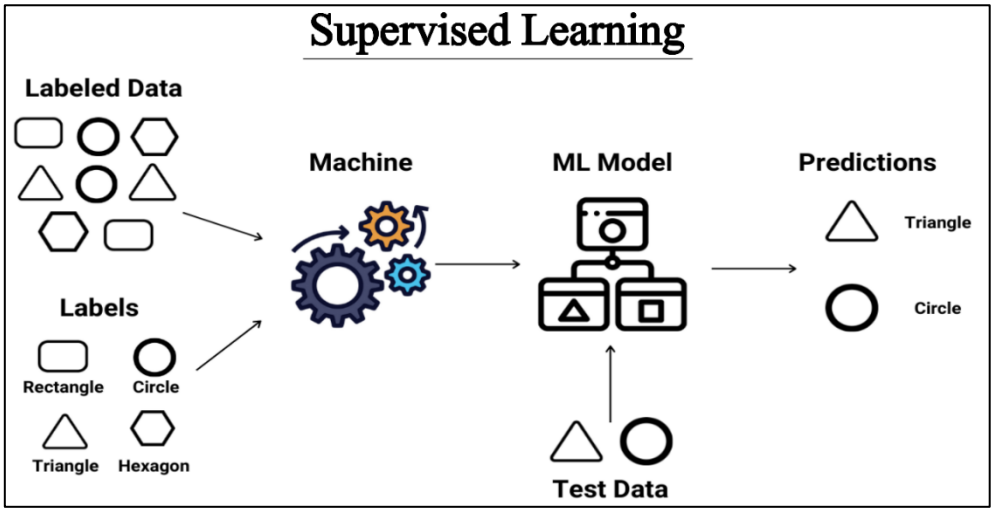


Figure 1.3. Conceptual scheme of supervised learning workflow (modified from Enjoy Algorithms).

1.2.3.2 Unsupervised Learning

Unsupervised learning deals with unlabeled data, seeking to uncover hidden patterns or structures within the data without predefined target variables (Figure 1.4). This approach is particularly useful when we don't know what patterns to look for in advance (Hastie et al., 2009). Common tasks in unsupervised learning include clustering (grouping similar data points), dimensionality reduction (reducing the number of features while preserving important information), and anomaly detection (identifying unusual data points). Algorithms like K-means clustering, principal component analysis (PCA), and autoencoders fall under this category (Murphy, 2012). Unsupervised learning is crucial in exploratory data analysis,

customer segmentation, and feature learning, providing insights that can guide further analysis or inform decision-making processes (Bishop, 2006).

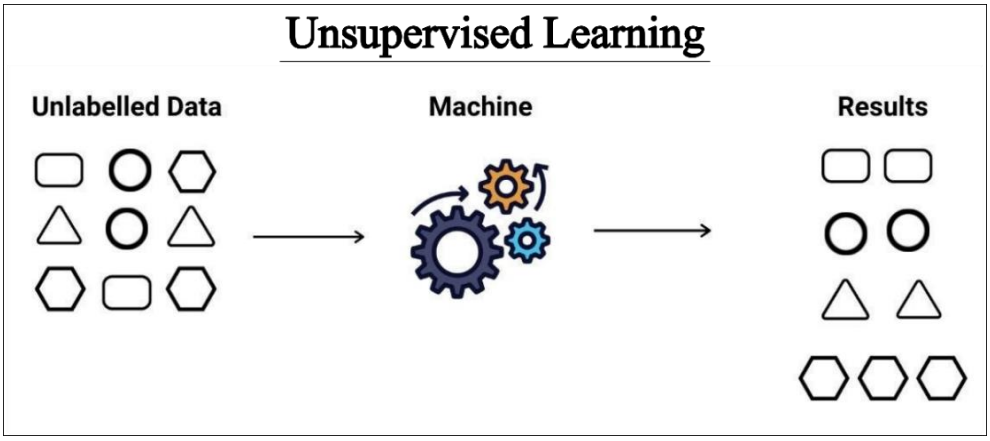


Figure 1.4. Conceptual unsupervised learning workflow scheme (modified from Enjoy Algorithms).

1.2.3.3 Semi-Supervised Learning

Semi-supervised learning bridges the gap between supervised and unsupervised learning, utilizing a small amount of labeled data along with a larger amount of unlabeled data (Figure 1.5). This approach is particularly valuable when obtaining labeled data is expensive, time-consuming, or requires expert knowledge (Zhu & Goldberg, 2009). By leveraging the structure in the unlabeled data, semi-supervised learning can often achieve better performance than using the labeled data alone. This makes it especially useful in domains like web content classification, speech analysis, and protein sequence classification, where vast amounts of unlabeled data are available, but labeling is resource-intensive (Chapelle et al., 2006).

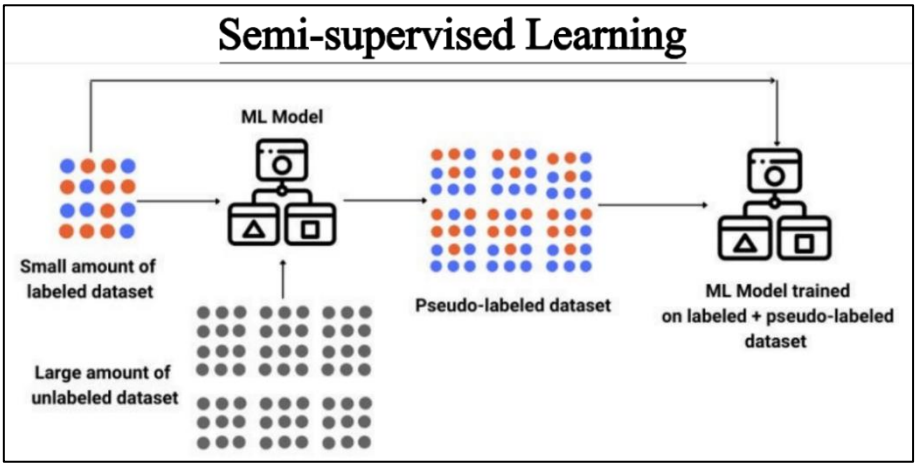


Figure 1.5. Conceptual semi-supervised learning scheme (modified from Enjoy Algorithms).

1.2.3.4 Reinforcement Learning

Reinforcement learning (RL) involves an agent learning to make decisions by interacting with an environment (Figure 1.6). Unlike supervised learning, there are no labeled examples of correct behavior. Instead, the agent learns to take actions that maximize a cumulative reward signal (Sutton & Barto, 2018). This approach involves a delicate balance between exploration (trying new actions to gather information) and exploitation (using known information to maximize reward). RL has gained significant attention due to its success in complex decision-making tasks, such as game playing (e.g., AlphaGo beating world champions in Go) and robotics (Silver et al., 2016). It's particularly suited to problems where long-term strategy is important and the optimal sequence of decisions is not immediately apparent (Sutton & Barto, 2018).

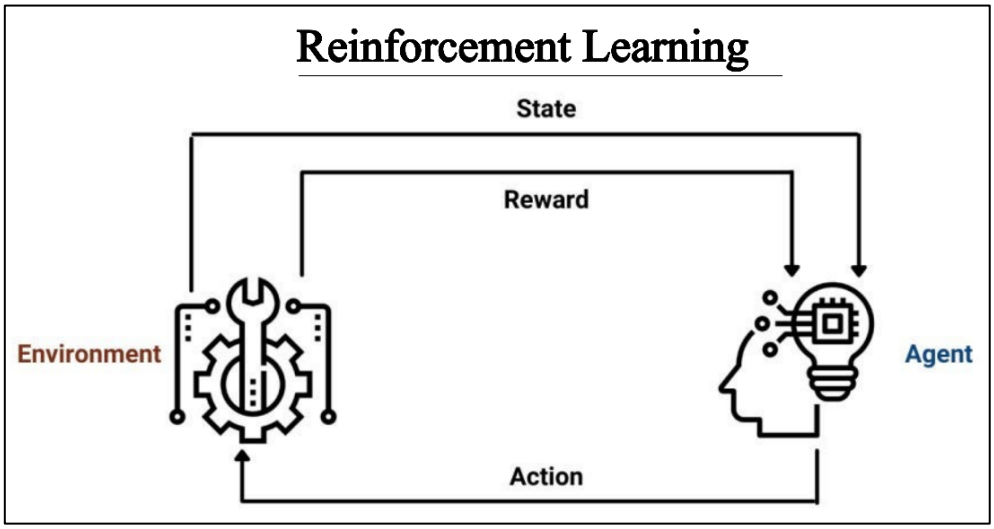


Figure 1.6. Conceptual reinforcement learning scheme (modified from Enjoy Algorithms).

1.2.3.5 Transfer Learning

Transfer learning is a valuable technique in machine learning that enables models to apply knowledge gained from one task to improve performance on another related task. As illustrated in Figure 1.7, a convolutional neural network (CNN) might first be trained from scratch on a large dataset of images, such as those of cars, trucks, and bicycles, where it learns important features. Once this pre-trained model is established, it can be adapted for a new task, such as classifying vehicles. Instead of starting from scratch, the model uses the knowledge it has already acquired, allowing it to be fine-tuned with a smaller dataset of vehicle images. This process not only saves time and computational resources but also enhances accuracy, as the model can achieve better results with fewer examples. Essentially, transfer learning makes it easier to develop effective models, especially in situations where collecting large amounts of data for the new task is challenging (Pan & Yang, 2010; Goodfellow et al., 2016).

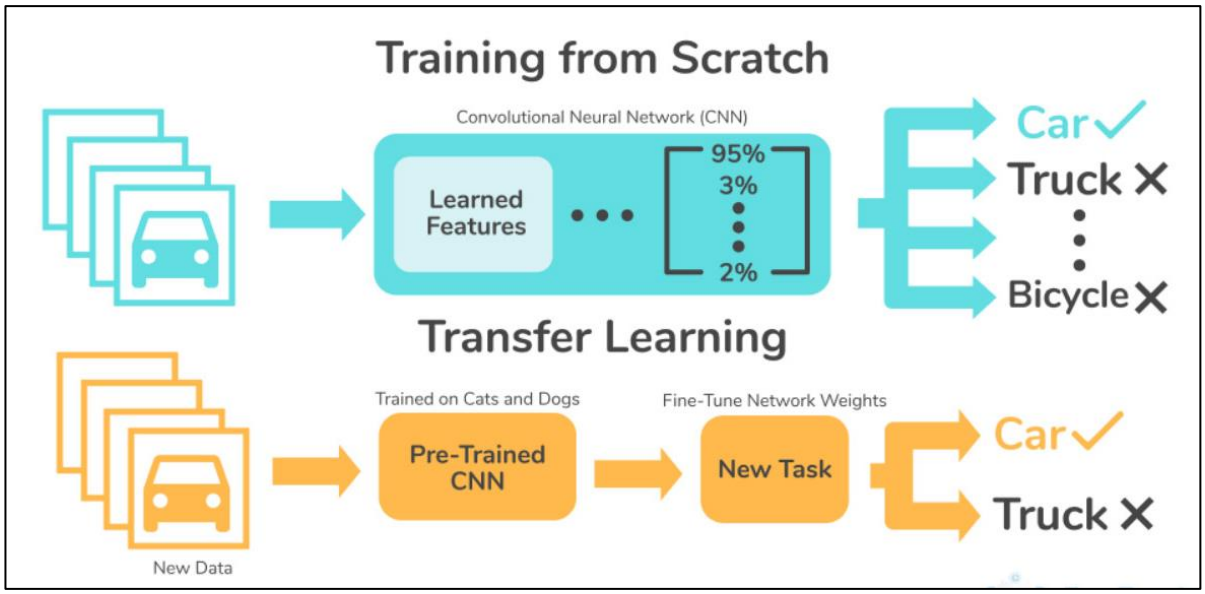


Figure 1.7. Conceptual examples on different workflows between a CNN and a transfer learning algorithm, (modified from Sky Engine).

1.2.4 Foundational Algorithms

This section delves into the fundamental algorithms and techniques that form the bedrock of machine learning. These methods, while sometimes considered "classical" in the rapidly evolving field of ML, remain crucial for understanding more advanced concepts and are still widely used in practice.

1.2.4.1 Linear Regression

Linear regression is one of the oldest and most interpretable algorithms in machine learning. It models the relationship between input variables (features) and a continuous output variable by fitting a linear equation to observed data (Seber & Lee, 2012).

The basic form of linear regression can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad [1.1]$$

Where y is the target variable, $x_1, x_2, \ldots, x_n$ are the input features, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients, and ε is the error term (Draper & Smith, 1998).

The algorithm aims to find the values of β that minimize the sum of squared residuals. This is typically done using methods like ordinary least squares or gradient descent (Murphy, 2012).

Linear regression's strengths lie in its simplicity, interpretability, and computational efficiency. It's widely used in economics for trend analysis, in finance for portfolio management, and in various scientific fields for understanding relationships between variables (Kutner et al., 2004).

However, it assumes a linear relationship between variables and is sensitive to outliers (Figure 1.8). It may underperform on complex, non-linear relationships in data (Hastie et al., 2009).

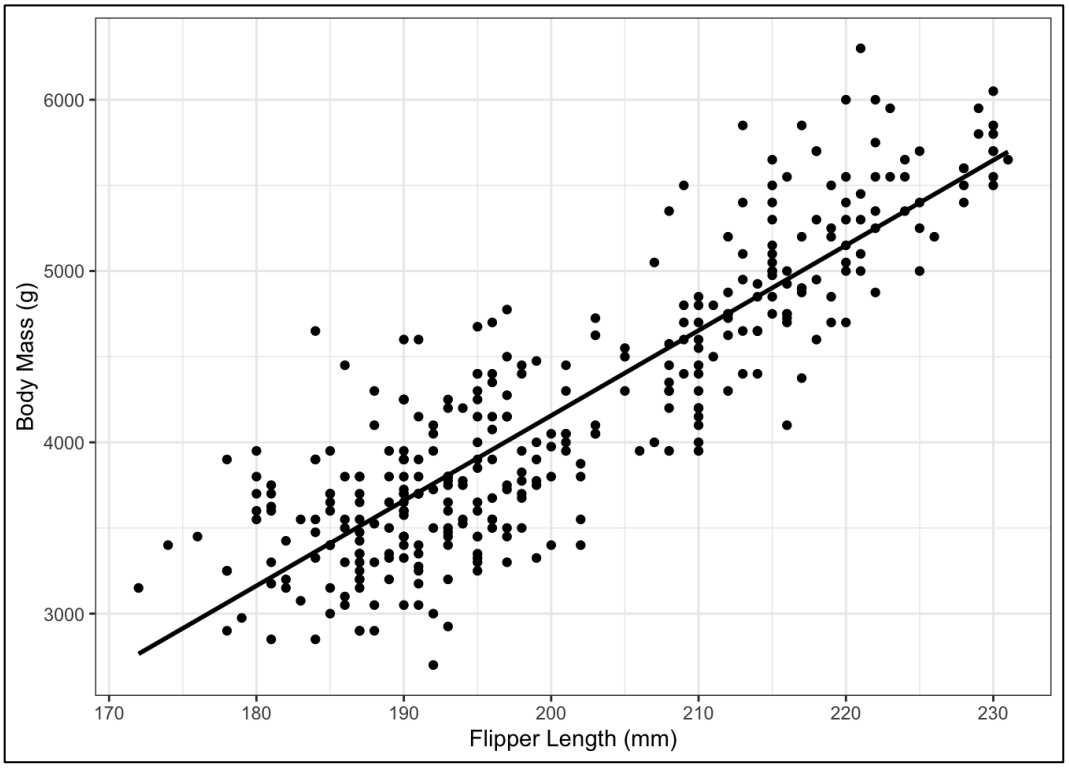


Figure 1.8. Simple linear regression example (modified from Six Sigma).

1.2.4.2 Logistic Regression

Despite its name, logistic regression is primarily used for binary classification tasks. It predicts the probability of an instance belonging to a particular class by applying the logistic function to a linear combination of features (Figure 1.9) (Hosmer et al., 2013).

The logistic function (sigmoid) is defined as:

$$\sigma(z) = \frac{1}{(1 + e^{(-z)})} \qquad [1.2]$$

Where z is the linear combination of features: $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$

Logistic regression seeks to determine the optimal values of β to maximize the likelihood of the observed data, typically achieved through maximum likelihood estimation or gradient descent algorithms (Murphy, 2012). Its ability to provide probabilistic outputs makes it particularly valuable in applications like risk assessment in credit scoring, medical diagnosis,

and customer behavior prediction in marketing (Menard, 2002). However, a key limitation of logistic regression is its assumption of linearity in the log-odds space, which may not accurately capture complex relationships within the data (Bishop, 2006).

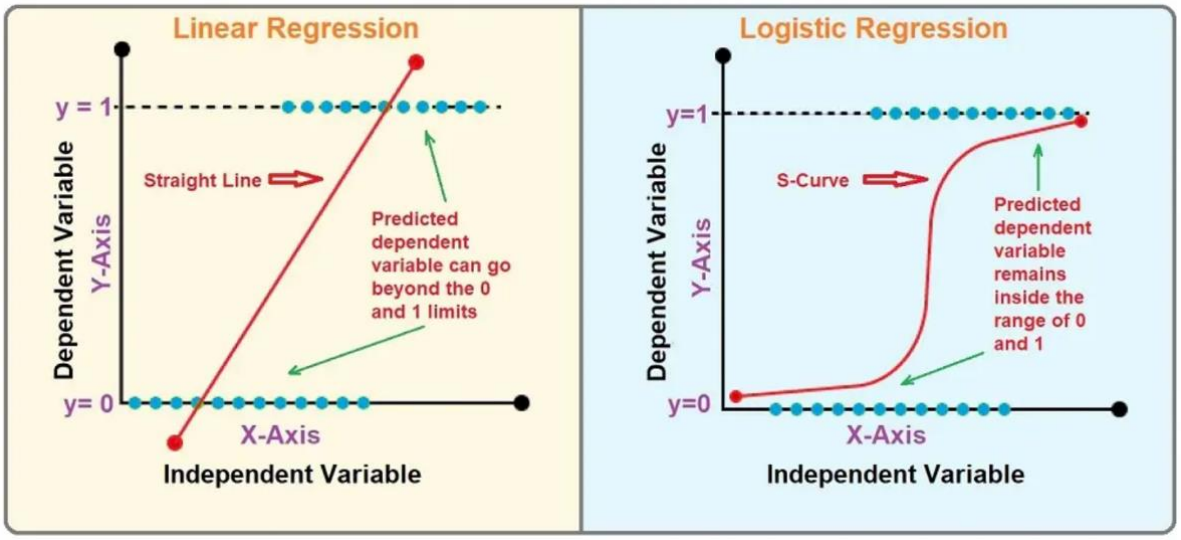


Figure 1.9. Scheme showing differences between Linear Regression and Logistic Regression logistic regression example (modified from Ejable).

1.2.4.3 Decision Trees

Decision trees are versatile machine learning algorithms used for both classification and regression tasks. They create a predictive model by learning simple decision rules from data features (Quinlan, 1986). The structure of a decision tree resembles a flowchart, starting from a root node and branching out based on feature-specific questions. This process continues until reaching leaf nodes that provide final predictions. For example, a tree predicting whether to play tennis might first ask about rain, then temperature, and so on (Figure 1.10).

Tree construction follows a top-down approach, recursively selecting the best feature to split the data by maximizing information gain or minimizing impurity (Breiman et al., 1984). This process continues until a stopping criterion is met, such as maximum tree depth. Decision trees offer several advantages: they are highly interpretable, handle both numerical and categorical data, and perform automatic feature selection. They effectively capture non-linear relationships and feature interactions (Loh, 2011). However, they can overfit if grown too deep, which can be mitigated by techniques like pruning or ensemble methods such as Random Forests (Breiman, 2001).

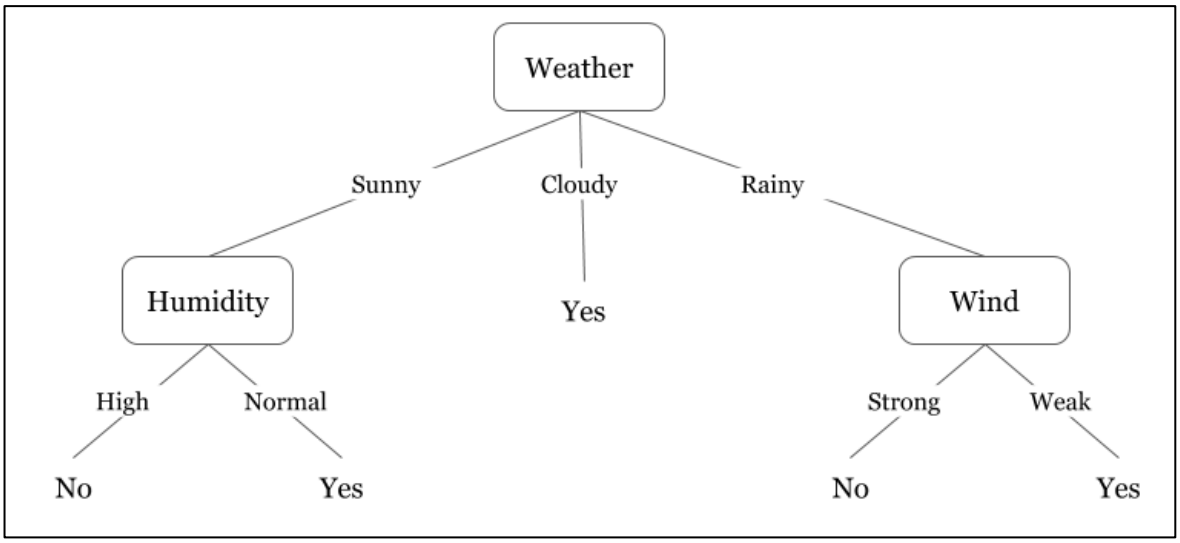


Figure 1.10. An understandable visualization of a decision tree scheme for playing tennis, (modified from Hacker Earth).

1.2.4.4 Support Vector Machines (SVM)

Support Vector Machines (SVM) are effective algorithms used primarily for classification and regression tasks. SVMs are known for their robust performance and strong theoretical foundation (Cortes and Vapnik, 1995). The core concept behind SVMs is to find the optimal hyperplane that best separates different classes in a high-dimensional space. For linearly separable data, SVM identifies the hyperplane that maximizes the margin between classes, which is defined as the distance from the hyperplane to the nearest data points of each class, known as support vectors (Figure 1.11).

In situations where data cannot be separated linearly, SVM uses the "kernel trick" to map the data into a higher-dimensional space where it becomes linearly separable. Common kernel functions include linear, polynomial, and radial basis function (RBF) (Schölkopf and Smola, 2002). The decision function for SVM classification can be expressed as:

$$f(x) = sign(\Sigma_i \alpha_i y_i K(x, x_i) + b) \quad [1.3]$$

Where α_i are the Lagrange multipliers, y_i are the class labels, K is the kernel function, x_i are the support vectors, and b is the bias term.

SVMs are particularly effective in high-dimensional spaces, especially when the number of features exceeds the number of samples. They offer a clear margin of separation, which often leads to better generalization on unseen data. SVMs are widely used in applications like text classification, image recognition, and bioinformatics (Ben-Hur et al., 2008). However, SVMs have limitations, such as not providing probability estimates directly and being sensitive to the

choice of kernel and hyperparameters. Additionally, training can be computationally expensive for very large datasets.

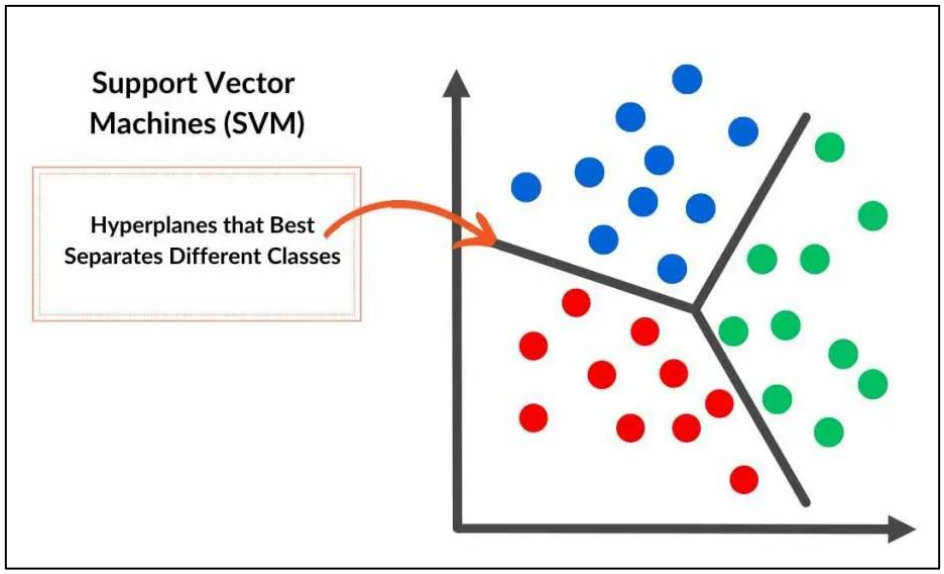


Figure 1.11. SMV conceptual scheme (modified from Spot Intelligence).

1.2.4.5 K-Means Clustering

K-means clustering is a fundamental algorithm used for unsupervised learning, particularly for data partitioning and pattern recognition. It partitions *n* observations into *k* clusters, with each observation assigned to the cluster with the nearest mean, or centroid (MacQueen, 1967).

The algorithm operates iteratively: it starts by initializing *k* centroids randomly, assigns each data point to the nearest centroid, recalculates the centroids based on the assigned points, and repeats these steps until convergence or a maximum number of iterations is reached (Figure 1.12). This process minimizes the within-cluster sum of squares (WCSS), defined as (Hartigan and Wong, 1979):

$$WCSS = \Sigma_i \Sigma_x ||x - \mu_i||^2 \quad [1.4]$$

Where x is a data point and μ_i is the centroid of cluster i.

K-means is popular across various domains for its simplicity and scalability, finding applications in customer segmentation, image compression, and anomaly detection (Jain, 2010). Its efficiency with large datasets makes it particularly valuable in the age of big data. However, K-means has limitations: it requires the number of clusters *k* to be specified in advance, which can be difficult when the data structure is unknown. The algorithm is also sensitive to initial centroid placement and assumes spherical clusters, which may not always match the data distribution (Arthur and Vassilvitskii, 2007). Despite these drawbacks, K-means

remains a powerful tool in machine learning, providing a foundation for more complex clustering methods and continuing to be widely used in data analysis.

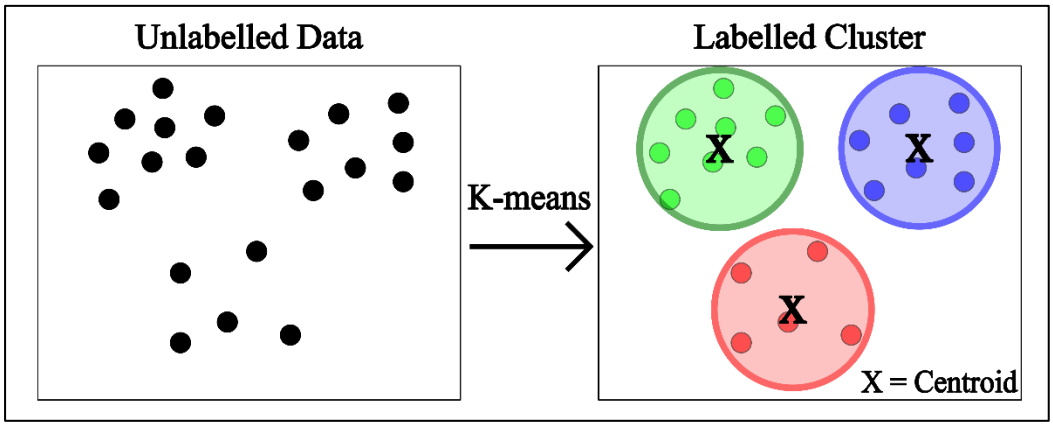


Figure 1.12. K-means scheme showing cluster labelling.

1.2.5 Deep Learning: A Paradigm Shift

Deep Learning has revolutionized the field of Machine Learning, driving substantial progress in diverse areas such as computer vision, natural language processing, and speech recognition. This paradigm shift is characterized by the use of neural networks with multiple layers, capable of learning hierarchical representations from data (LeCun et al., 2015).

The core of deep learning comprises several key architectures: Neural Networks (NNs), the fundamental building blocks; Convolutional Neural Networks (CNNs), specialized for processing grid-like data such as images; Recurrent Neural Networks (RNNs), designed for sequential data; and Long Short-Term Memory (LSTM) networks, which address the long-term dependency problems in standard RNNs.

The success of deep learning can be attributed to three main factors: the availability of large datasets, increased computational power, and algorithmic innovations. These elements have synergistically enabled the training of complex models that often surpass human-level performance on specific tasks (Goodfellow et al., 2016).

1.2.5.1 Neural Networks (NNs)

Neural Networks (NNs) form the foundation of deep learning, inspired by the neural structure of the human brain. These networks consist of interconnected layers of nodes, or neurons, that process and transmit information (Rosenblatt, 1958).

The simplest form of a neural network is the feedforward network, where data flows unidirectionally from input to output through one or more hidden layers. Each neuron applies a

non-linear activation function to its inputs, with popular choices including the Rectified Linear Unit (ReLU) for its ability to mitigate the vanishing gradient problem (Nair and Hinton, 2010).

Training a neural network involves adjusting the weights of connections between neurons to minimize the error between predicted and actual outputs (LeCun et al., 1998) (Figure 1.13). This is typically achieved through backpropagation, a method that efficiently computes gradients in neural networks by propagating the error backwards from the output to the input layers (Rumelhart et al., 1986). The refinement of this algorithm (LeCun et al. 1998) played a crucial role in reigniting interest in neural networks. Additionally, the introduction of boosting algorithms, particularly AdaBoost (Freund & Schapire, 1997), which iteratively combines weak learners to create a strong predictor by focusing on misclassified examples, further advanced the field of machine learning and complemented neural network approaches. The strength of neural networks lies in their ability to approximate complex functions (Cybenko, 1989; Hornik, 1991).

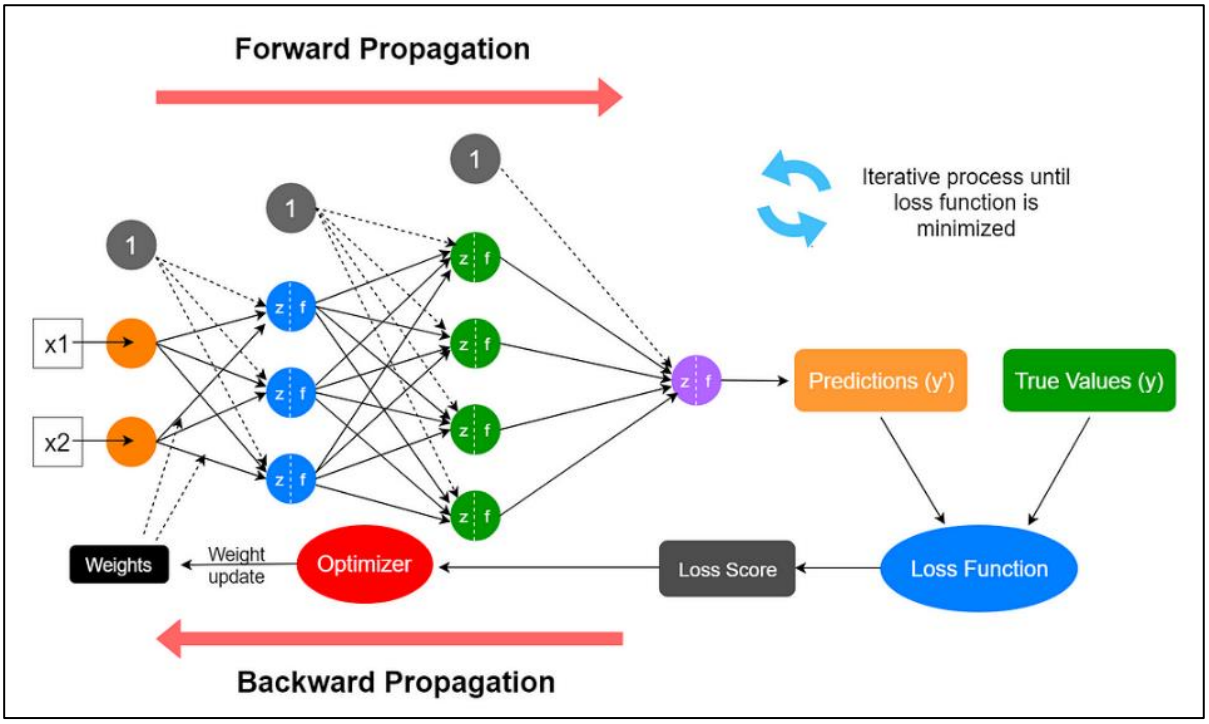


Figure 1.13. NNs conceptual workflow scheme (modified from Medium).

1.2.5.2 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are specialized neural networks designed to process sequential data, making them particularly suitable for tasks involving time series, speech, and natural language (Elman, 1990). The key feature of RNNs is their ability to maintain a hidden state that captures information from previous time steps, enabling them to model temporal dependencies in data (Figure 1.14).

In RNNs, the hidden state is updated at each time step based on the current input and the previous hidden state. This recurrent structure allows the network to retain a form of memory, making it well-suited for tasks where the order of data is crucial (Mikolov et al., 2010).

However, traditional RNNs face challenges with long-term dependencies due to the vanishing and exploding gradient problems, which can make training difficult for long sequences (Bengio et al., 1994). These limitations led to the development of more advanced RNN architectures, most notably the Long Short-Term Memory (LSTM) networks.

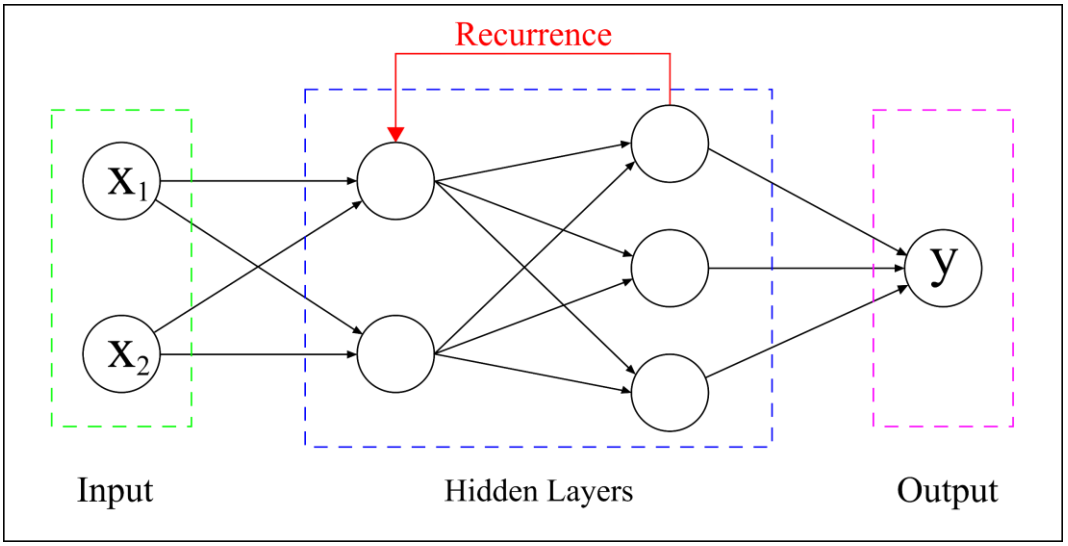


Figure 1.14. RNN conceptual workflow scheme.

1.2.5.3 Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks are a specialized type of Recurrent Neural Networks (RNNs) designed to overcome the limitations of traditional RNNs in capturing long-term dependencies. Introduced in 1997, LSTMs have become essential for sequence modeling tasks that require retaining information over extended periods (Hochreiter & Schmidhuber, 1997).

The key innovation of LSTMs is the memory cell, which maintains its state over time. This cell is managed by three gates: the input gate controls what new information to store, the forget gate decides what information to discard, and the output gate determines what information should be output. This gating mechanism allows LSTMs to selectively remember or forget information, making them effective at capturing long-term dependencies in sequential data.

LSTMs have achieved significant success in various applications. In natural language processing, they have advanced machine translation, text generation, and sentiment analysis (Sutskever et al., 2014). In time series analysis, they excel in tasks such as stock price prediction and weather forecasting (Graves, 2012).

Despite their effectiveness, LSTMs come with challenges. They can be computationally expensive to train, particularly on long sequences, and often require large amounts of data to perform optimally. While they help mitigate the vanishing gradient problem, they can still struggle with very long-term dependencies.

As research in deep learning progresses, variants like Gated Recurrent Units (GRUs) have been proposed to address some limitations while retaining the core benefits of the LSTM architecture (Cho et al., 2014).

1.2.5.4 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are specialized neural networks designed for processing grid-like data, particularly images. Inspired by the organization of the animal visual cortex, CNNs have transformed computer vision by automatically learning hierarchical features from raw image data (LeCun et al., 1998). The training process of a CNN consists of two main phases: feedforward and backpropagation, as illustrated in Figure 1.15.

During the feedforward phase, input data is processed through multiple layers, including convolutional layers that apply learnable filters (or kernels) to detect local patterns like edges and textures. This generates feature maps that highlight specific characteristics at various spatial locations (Krizhevsky et al., 2012). Following convolution, pooling layers, typically max pooling, reduce the spatial dimensions of these feature maps. This downsampling retains essential information while decreasing computational complexity, allowing CNNs to progressively learn more abstract features as the network deepens (Zeiler and Fergus, 2014).

The backpropagation phase follows, where the model's predictions are compared to actual target values, and the loss is calculated. This loss guides the adjustment of weights in the network, refining the model's ability to make accurate predictions. The final layers of a CNN, usually fully connected layers, perform high-level reasoning based on the extracted features, combining them in optimal ways for specific tasks such as classification, detection, or segmentation.

CNNs have shown remarkable success across various computer vision tasks, consistently outperforming traditional methods. Architectures like AlexNet, VGGNet, and ResNet have achieved significant results on challenging datasets like ImageNet (He et al., 2016). Beyond image classification, CNNs are widely applied in object detection (e.g., YOLO, SSD) and image segmentation (e.g., U-Net), with applications spanning medical image analysis, autonomous driving, and satellite imagery interpretation. Their influence extends to video analysis, face recognition, and non-visual domains such as speech recognition and natural language processing. As CNNs continue to evolve, researchers explore ways to enhance their efficiency

and interpretability, addressing challenges like computational complexity and improving generalization to out-of-distribution samples.

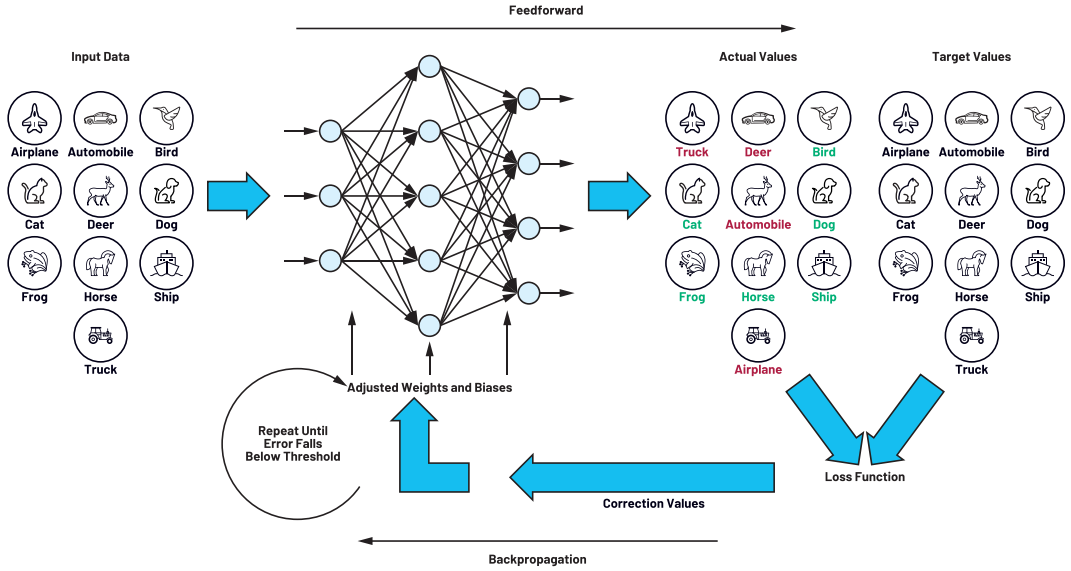


Figure 1.15. Conceptual scheme of a Convolutional Neural Network (CNN) training loop, illustrating the feedforward and backpropagation processes (modified from Analog).

1.2.6 Current State of the Art in Machine Learning

1.2.6.1 Transformers

Transformer models have transformed natural language processing (NLP) and have significantly influenced other fields. Transformers change how machine learning models handle sequences of data. Unlike Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), transformers can process entire data sequences at once due to their self-attention mechanism (Vaswani et al., 2017). This capability helps them effectively capture long-range dependencies, making them especially useful for tasks like understanding and generating language (Vaswani et al., 2017).

The self-attention mechanism is the heart of transformers. It analyzes relationships between different parts of an input sequence using three components: Query (Q), Key (K), and Value (V). This approach allows the model to determine the importance of each element in the sequence, regardless of its position. Additionally, multi-head attention lets transformers identify multiple relationships within the data simultaneously, enhancing their ability to process complex inputs (Figure 1.16).

The transformer architecture consists of an encoder and decoder, each made up of identical layers that incorporate self-attention and fully connected feed-forward networks. This design allows the model to focus on relevant parts of the input when generating output, making it effective for various sequential tasks.

Their impact has led to innovations in various domains: Several transformer-based models have significantly advanced NLP. BERT (Bidirectional Encoder Representations from Transformers) excels in tasks like question answering and sentiment analysis by understanding word context in both directions (Devlin et al., 2019). GPT (Generative Pre-trained Transformer), developed by OpenAI, focuses on language generation and has shown impressive abilities in translation, text completion, and creative writing. The latest versions, like GPT-3, demonstrate few-shot learning capabilities, allowing them to perform well with minimal task-specific fine-tuning (Radford et al., 2018; Brown et al., 2020).

While initially developed for NLP, transformers have applications beyond text processing:

- Computer Vision: Vision Transformers (ViTs) have been applied to image classification, object detection, and segmentation tasks (Dosovitskiy et al., 2021). Treating image patches as tokens enables effective visual data processing, with models like SegFormer achieving state-of-the-art results in semantic segmentation (Xie et al., 2021).
- Reinforcement Learning: Transformers are utilized in reinforcement learning tasks, helping agents learn from sequences of actions and states in environments like games and robotics (Chen et al., 2021).

The rise of transformer models marks a paradigm shift in machine learning, initially transforming NLP and inspiring advancements across multiple fields. The transformer architecture has proven to be a powerful framework for processing various types of sequential and structured data, setting new benchmarks in language tasks and inspiring similar applications in other domains.

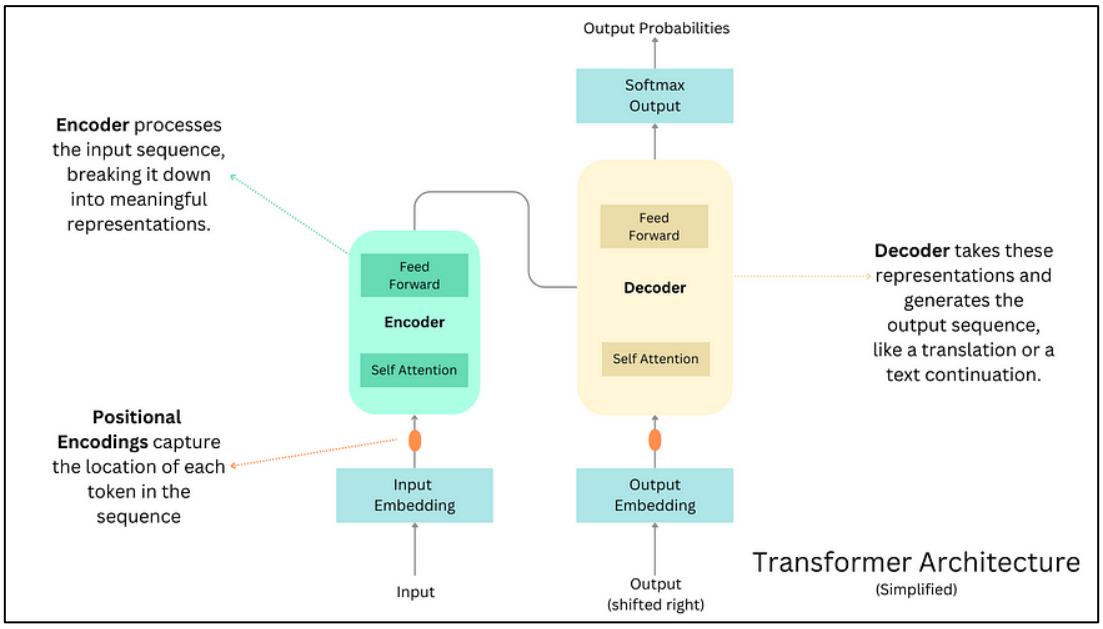


Figure 1.16 Transformer Architecture explained and simplified (modified from Medium).

1.2.6.2 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs), represent another significant advancement in machine learning, particularly in the domain of generative modeling (Goodfellow et al., 2014). GANs consist of two neural networks, a generator and a discriminator, that are trained simultaneously through an adversarial process, where the generator attempts to create data (e.g., images) that resemble the training set, while the discriminator tries to distinguish between real and generated data. This adversarial training leads to the generator producing increasingly realistic outputs as it learns to fool the discriminator (Fig. 1.17). Architecture's ability to generate highly realistic synthetic data has made GANs a powerful tool in various domains (Goodfellow et al., 2014; Radford et al., 2016).

GANs have found wide-ranging applications across various domains:

- Image Generation: Creating realistic images, including faces, artwork, and scene completions.
- Image-to-Image Translation: Converting images from one domain to another (e.g., sketches to photos, summer to winter scenes).
- Super-Resolution: Enhancing the resolution and quality of images.
- Data Augmentation: Generating synthetic data to enhance training datasets in fields like medical imaging.
- Video Generation: Creating realistic video sequences from static images or textual descriptions.

Despite challenges such as training instability and mode collapse, GANs continue to be at the forefront of generative modeling. Their ability to produce highly realistic synthetic data has opened up new possibilities in artificial data generation and manipulation, making them a crucial tool in modern machine learning applications.

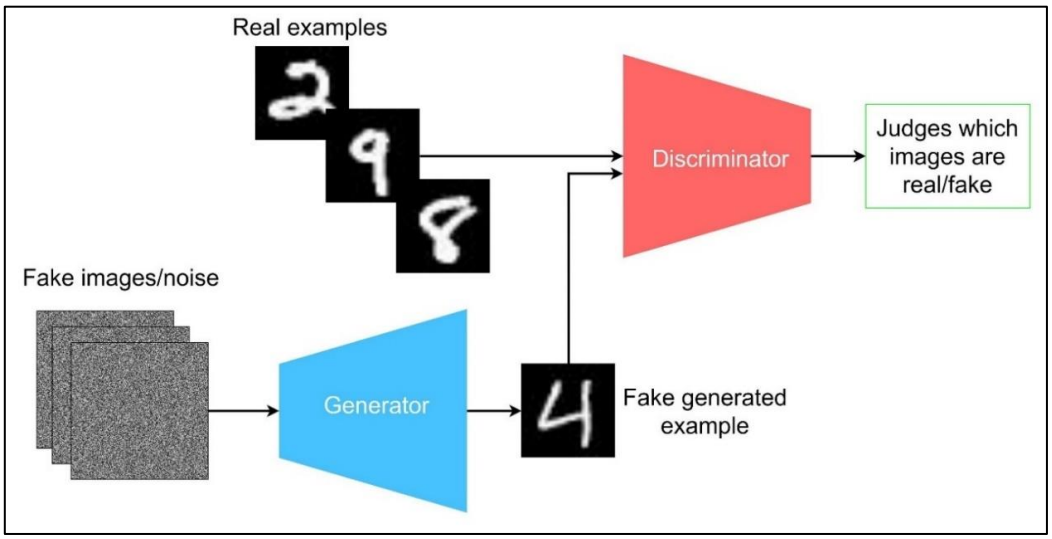


Figure 1.17. Simple scheme of GANs workflow (modified from IBM Developer).

1.2.7 Cutting-Edge Technologies in Machine Learning

The field of Machine Learning (ML) continues to advance rapidly, with cutting-edge technologies and research areas pushing the boundaries of what is possible. This section examines some of the most exciting and influential areas in contemporary ML, including few-shot and zero-shot learning, generative models, explainable AI, and federated learning.

1.2.7.1 Few-Shot and Zero-Shot Learning

Traditional machine learning models typically require large amounts of labeled data to achieve high performance. However, few-shot and zero-shot learning approaches aim to overcome this limitation by enabling models to generalize from very limited examples or even to perform tasks they have never explicitly been trained on (Figure 1.18).

Few-shot learning involves training models to recognize new classes or tasks with only a few examples. Techniques such as meta-learning, where the model learns to learn by using its training experience across many tasks, have been particularly effective (Finn et al., 2017).

Zero-shot learning takes this concept a step further, enabling models to perform tasks without any direct examples from those tasks. This is often achieved by leveraging semantic information from other tasks or domains. For instance, a model might use word embeddings to infer relationships between known and unknown classes, allowing it to generalize knowledge from seen to unseen classes (Xian et al., 2018).

These approaches are particularly useful in scenarios where data collection is expensive or impractical, such as rare disease diagnosis or real-time personalization in recommendation systems. The potential of few-shot and zero-shot learning to dramatically reduce the amount of labeled data required for effective model training could make ML more accessible and applicable in a wider range of scenarios.

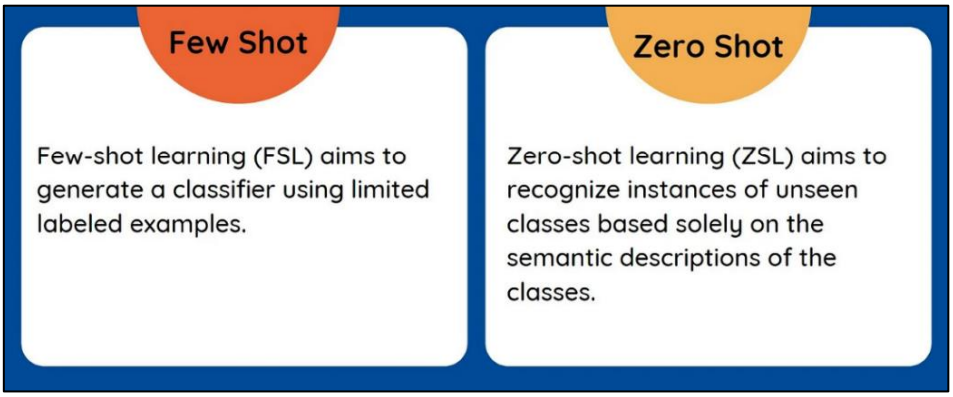


Figure 1.18. Few-shot learning compared with Zero-shot learning simplified (modified from Medium).

1.2.7.2 Explainable AI (XAI)

As machine learning (ML) models become more complex and widespread, the need for explainability and interpretability has grown significantly. Explainable AI (XAI) aims to make the decision-making processes of ML models transparent and understandable (Gunning & Aha, 2019).

Several methods have been developed to enhance explainability:

- Model-agnostic techniques, like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations), provide individual predictions by approximating complex models with simpler, interpretable ones (Ribeiro et al., 2016; Lundberg & Lee, 2017).
- Efforts to create interpretable versions of complex models, such as neural networks, strive to maintain performance while offering clearer insights into their decision-making (Rudin, 2019).
- Visualization tools, including saliency maps and attention visualizations, highlight significant parts of input data in a model's predictions, providing intuitive insights into model behavior (Simonyan et al., 2013).

XAI is essential for building trust in ML systems, particularly in critical fields. As ML systems increasingly influence important decisions, the ability to explain and interpret these systems will be vital for their acceptance and adoption (Adadi & Berrada, 2018). The XAI field is evolving to tackle current technical and ethical challenges, promoting responsible and inclusive use of ML technologies (Figure 1.19). By enhancing transparency and accountability, XAI techniques are shaping the future of AI and its societal impact (Arrieta et al., 2020).

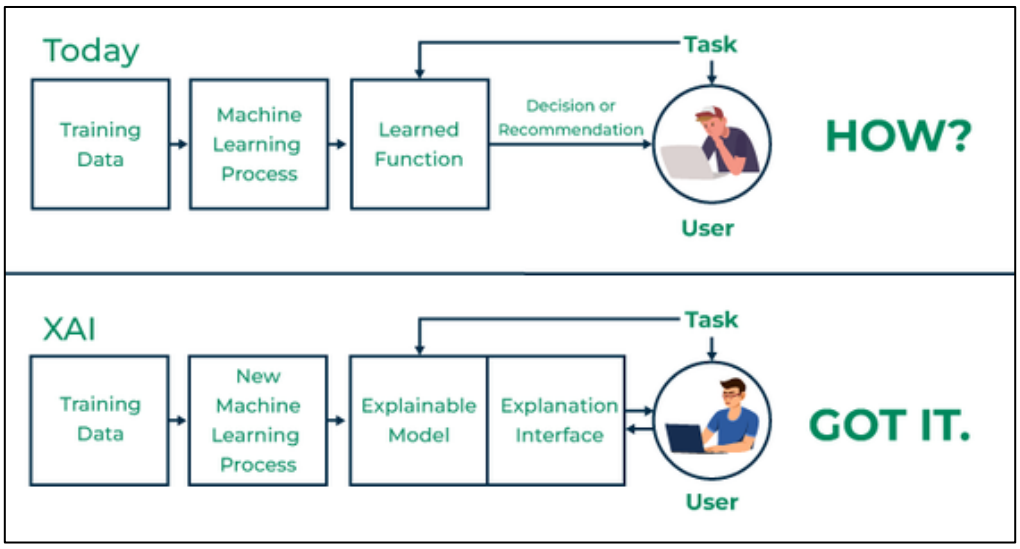


Figure 1.19. Explainable Artificial Intelligence (XAI) conceptual example (modified from Geeks for geeks).

1.3 Machine Learning Real Use Cases for Civil Protection

1.3.1 Prediction of Rainfall-Induced Landslides with Deep Learning

The prediction of rainfall-induced landslides is a critical aspect of disaster management, particularly in regions prone to heavy rainfall and complex topography. Traditional methods for landslide prediction often rely on empirical rainfall thresholds or physically-based models that require detailed environmental data (Guzzetti et al., 2008; Segoni et al., 2018). However, the advent of machine learning (ML) techniques has opened new avenues for more accurate and scalable prediction models. These models can handle large datasets, learn complex patterns in the data, and provide timely predictions, which are essential for early warning systems and disaster mitigation (Piciullo et al., 2018).

Deep learning models, especially deep learning approaches, offer significant advantages in predicting landslides triggered by rainfall. These models can analyze vast amounts of historical rainfall and landslide data to identify patterns that precede landslide events. By learning from past events, they can provide real-time predictions, even in diverse and complex terrains (Mondini et al., 2023). Moreover, the ability to integrate various data sources, such as satellite rainfall estimates, ground-based observations, and topographical information, further improves their predictive capabilities (Kirschbaum and Stanley, 2018).

- **Zhao et al. (2022)** developed an AI-based rainfall prediction model for debris flows, focusing on a high-frequency debris flow catchment in central China. The model uses continuous rainfall data from five rain gauges, covering 367 rainfall events between 2012 and 2015, with 46 events triggering debris flows and 321 events without debris flows. A total of 17 machine learning algorithms were tested, and 141 rainfall features were extracted for the model training. The Extra Trees (ETs) model performed best, predicting debris flow events with high accuracy. During testing with simulated real-time rainfall data, the model showed no false alarms or missing alarms, predicting debris flows up to 35 minutes in advance. While promising, the results are based on simulations, highlighting the model's potential for early warning, though further real-world testing is needed for full validation.
- **Mondini et al. (2023)** developed a deep learning system to forecast rainfall-induced shallow landslides in Italy. The model uses hourly rainfall data from 2,096 gauges and information on 2,486 landslides (from 2002 to 2020). In total, the dataset includes 780,766 rainfall events (Fig. 1.20), of which 2,472 were associated with landslides, while the remaining 778,294 did not trigger any landslides. The system employs an ensemble of 2,400 neural network models to predict landslide occurrence based solely on rainfall history, without needing detailed terrain data. The model generates probabilistic forecasts for 1 to 24 hourly lag periods, providing a sharp dichotomous (landslide/no landslide) prediction for each hour. The model demonstrated high predictive accuracy (AUC 0.87–

0.91), proving effective for landslide forecasting and well-suited for integration into operational early warning systems across large regions.

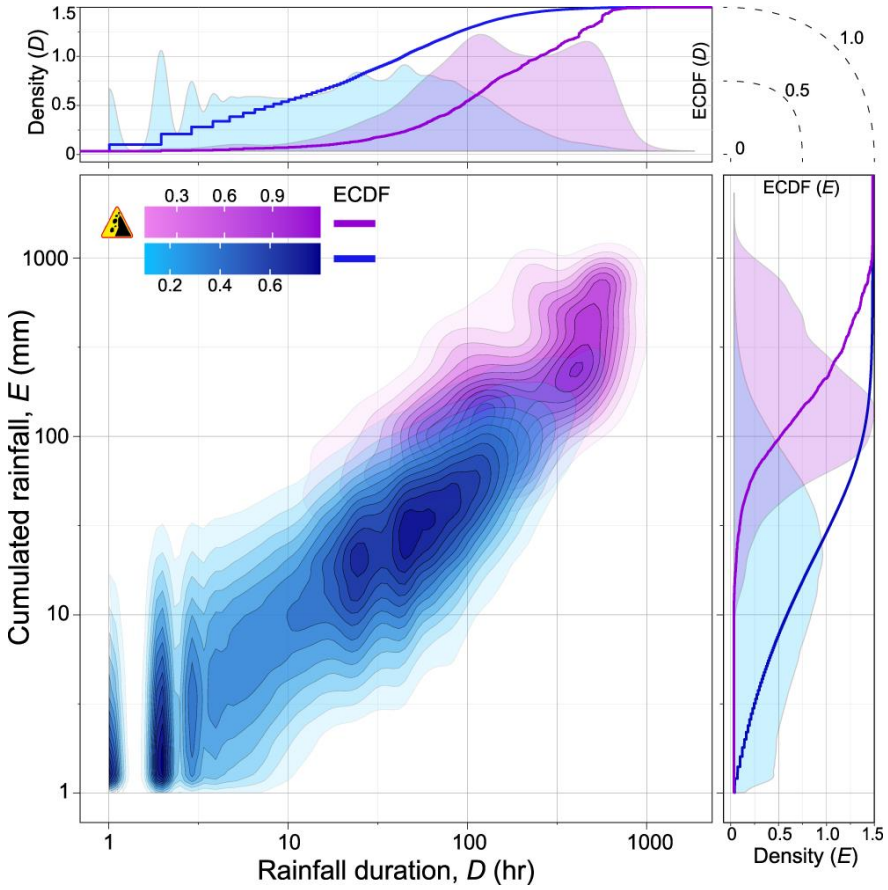


Figure 1.20. The figure shows the density of 2,472 rainfall events (in violet) from February 2002 to December 2020 that triggered at least one landslide, compared to 778,294 rainfall events (in blue) that did not result in any known landslides. The upper and right plots display the marginal distributions of duration (D) and cumulative rainfall (E), with violet and blue areas representing landslide-triggering and non-landslide-triggering events, respectively. The corresponding empirical cumulative distribution functions (ECDF) are normalized to the total number of rainfall events with and without landslides. Mondini et al. (2023).

- **Guzzetti et al. (2024)** present a deep-learning system to forecast rainfall-induced landslides across Italy, using an ensemble of 2,400 neural network models. The system incorporates rainfall data from 2,096 rain gauges and 2,486 landslides recorded between 2002 and 2020, with the CTRL-T software calculating key rainfall variables for each event. Forecasts are generated for 24 hourly lag periods, and a voting scheme classifies whether a landslide is likely based on the majority of predictions. The system demonstrated strong performance, particularly for events involving multiple landslides in the same area, achieving an 82.6% success rate in predicting landslide-related hourly records. Despite some challenges with more complex events like rockfalls, the authors

conclude that the system offers a valuable tool for short-term operational forecasting of rainfall-induced landslides, significantly enhancing emergency response efforts in Italy.

Despite the promising advancements in ML-based landslide prediction, a significant challenge remains in the collection and quality of input data. Landslide prediction models often require extensive, long-term datasets with relatively low uncertainty. However, in the field of landslide research, obtaining such data can be problematic for several reasons. Firstly, the collection of high-quality landslide data typically requires on-site instrumentation that can gather information over extended periods, often spanning years or decades (Angeli et al., 2000; Petley et al., 2005). This long-term data collection is both resource-intensive and logistically challenging, especially in remote or hazardous areas. Secondly, landslide triggering mechanisms are subject to a degree of interpretability and can involve complex, varied mechanical and kinematic processes (Iverson, 2000; Sidle and Ochiai, 2006). This inherent variability introduces uncertainty into the data, which can affect the performance of prediction models. Furthermore, the heterogeneity of landslide types and their local geological and hydrological contexts means that data collected in one area may not be fully applicable to another, limiting the generalizability of models (Crosta and Frattini, 2003; Guzzetti et al., 2007). While ML methods have shown great potential in improving landslide prediction accuracy, their effectiveness is ultimately constrained by the quantity, quality, and representativeness of the available data. Addressing these data-related challenges remains a crucial area for advancement in the field of landslide prediction and early warning systems (Baum and Godt, 2010; Guzzetti et al., 2020).

1.3.2 Landslide Mapping with Convolutional Neural Networks (CNNs)

The rapid and accurate mapping of landslides over large areas has long been a challenging task in the field of geology and disaster management. Traditional methods, relying on manual interpretation of satellite or aerial imagery, are time-consuming and often impractical in emergency scenarios. The advent of deep learning techniques, particularly Convolutional Neural Networks (CNNs), is revolutionizing the field, offering promising solutions for automated landslide detection and mapping.

In recent years, CNNs have achieved remarkable results in various fields of computer vision and image processing. Their success spans across a wide range of applications, including image classification (Dosovitskiy et al., 2021), object detection (Redmon et al., 2016), and semantic segmentation (Chen et al., 2017). Despite their widespread adoption in many domains, the application of CNNs to post-event landslide mapping has been relatively limited, with significant developments only emerging in the past few years.

- **Ding et al. (2016)** proposed an automatic landslide recognition method using Convolutional Neural Networks (CNNs) and texture change detection in Shenzhen, China.

They utilized GF-1 satellite images taken on December 25, 2013 (pre-landslide) and December 29, 2015 (post-landslide), with four spectral bands at an 8-meter resolution. The landslide areas were manually mapped to create a ground truth dataset. Their method involved three steps: removing irrelevant areas using spectral indices, extracting features with a CNN pre-trained on the SAT-6 dataset, and confirming landslide areas through texture change analysis. This approach demonstrated the effectiveness of integrating CNNs with traditional methods for improved landslide detection accuracy.

- **Ghorbanzadeh et al. (2019)** evaluated various methods for landslide detection in the Rasuwa district of Nepal. They used optical data from the RapidEye satellitewith a spatial resolution of 5 meters, alongside topographic factors. The dataset consisted of 3,500 original samples, which were augmented to approximately 7,000 image sample patches for training. They employed multiple machine learning techniques (ANN, SVM, RF), alongside CNNs with varying input window sizes and layer depths. The best results were achieved with a small window size CNN, yielding a mean intersection-over-union (mIOU) of 78.26%. The study revealing that while CNNs show promise, their performance is highly dependent on design choices and training strategies.

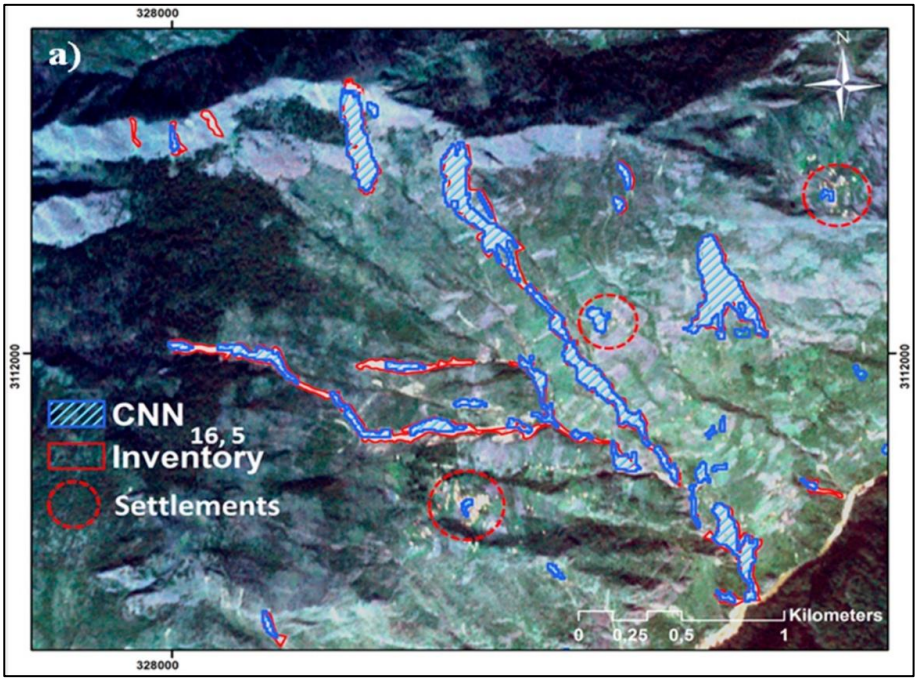


Figure 1.23. Comparison of landslide detection results between the automated mapping method and ground-truth data in the study area (Ghorbanzadeh et al. 2019, modified). The figure illustrates the performance of the algorithm in identifying landslides, showing both true positives (correctly identified landslides), false positives (areas incorrectly marked as landslides), and false negatives (missed landslide areas).

- **Meena et al. (2021)** proposed a deep learning approach for rapidly mapping landslides triggered by the 2018 monsoon rains in Karnataka, India. They utilized PlanetScope satellite imagery, which provides high-resolution data, and incorporated slope topographic data to create a robust training dataset. A total of 343 landslide polygons were manually delineated to establish ground truth. The study employed a fourfold cross-validation strategy to divide the data into training (75%) and testing (25%) sets, enhancing the model's robustness. Their structured CNN model, implemented in Trimble's eCognition software, used convolutional layers with kernel sizes varying from 5 × 5 to 3 × 3, combined with max pooling layers (2 × 2 kernel size). This configuration was optimized using a statistical gradient descent function, with a batch size of 50, learning rate of 0.0001, and 3000 epochs. The input window size of 32 × 32 pixels was chosen to capture the diverse shapes and sizes of landslides in the study area. Their CNN model showed an accuracy of 78% when including slope data, compared to 65.5% with optical data alone. This methodology effectively demonstrated the advantages of combining topographic features with deep learning for improved landslide detection (Fig. 1.24).

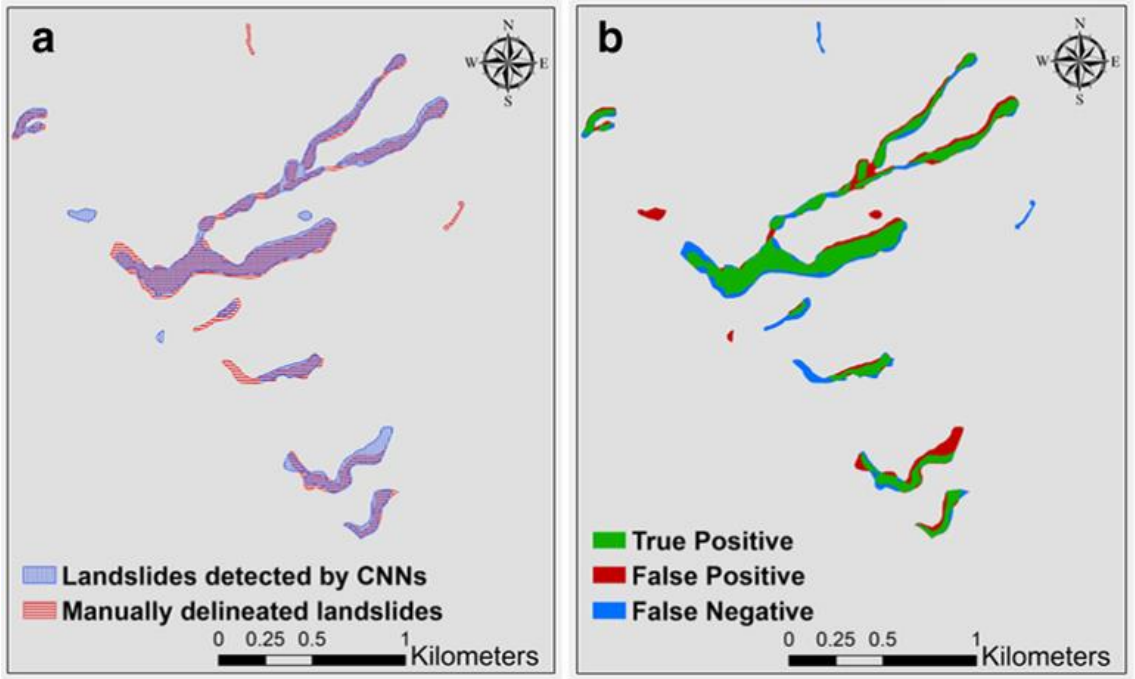


Figure 1.24. (**a**) Landslide areas cataloged through both CNN analysis and manual delineation. (**b**) Areas classified as true positives (TP), false positives (FP), and false negatives (FN) based on the spatial overlap analysis of the polygons identified in (a) (adapted from Meena et al. 2021, modified).

- **Ghorbanzadeh et al. (2022)** investigated landslide detection using an integration of deep learning and object-based image analysis (OBIA) within a study area affected by the Typhoon Morakot in Western Taiwan, which triggered nearly 22,700 landslides, manually mapped, across 274 km², manually mapped. They utilized Sentinel-2 satellite imagery,

with a spatial resolution of 10 meters. The methodology involved training a Residual U-Net (ResU-Net) model on five stacked layers of Sentinel-2 imagery, combined with a slope layer derived from ALOS. The ResU-Net's outputs were then refined using a simple rule-based OBIA approach. The study achieved significant improvements in landslide detection, highlighting the effectiveness of combining deep learning techniques with object-based analysis.

- **Nava et al. (2022)** focused on rapid landslide mapping in the eastern Iburi sub-prefecture of Hokkaido, Japan, following the Mw 6.6 earthquake that triggered around 7,837 co-seismic landslides, manually mapped from PlanetScope imagery. They utilized Sentinel-1 SAR amplitude imagery with a spatial resolution of 10 meters, applying a modified Attention U-Net model for pixel-based classification. The dataset was divided into training (70%) and testing (30%) sets to validate the model's performance. The best results were achieved using tri-temporal SAR data, yielding an F1-score of 61%.
- **Meena et al. (2023)** introduced the High-Resolution Global Landslide Detector Database (HR-GLDD), utilizing PlanetScope satellite imagery with a spatial resolution of 3 meters for landslide mapping across ten geographically diverse study sites affected by earthquake- and rainfall-triggered landslides. The dataset includes various landslide events and was systematically divided into training (60%), validation (20%), and testing (20%) sets, generating a total of 10,000 image patches. The models were evaluated using five advanced U-Net-like architectures. Their findings demonstrate the robustness and transferability of the dataset in detecting landslides across different topographies, including recent events in Haiti (2021), Indonesia (2022), and the Democratic Republic of the Congo (2020). This dataset marks a significant advancement in developing scalable and accurate models for rapid landslide detection worldwide.
- **Wu et al. (2024)** developed the SCDUNet++ model for landslide mapping, focusing on the Jiuzhaigou and Luding areas in China, which were impacted by the 2017 Sichuan earthquake. They utilized Sentinel-2 multispectral imagery with a spatial resolution of 10 meters, incorporating 13 bands along with four spectral indices (SIFs) and six topographic features (TFs). The dataset comprised a total of 1,840 images, split into training (60%), validation (20%), and testing (20%) sets. The model was validated using a combination of deep learning techniques, particularly CNNs, and demonstrated a notable performance with an F1-score of 85.68%. This study highlights the effectiveness of using advanced deep learning architectures for rapid landslide mapping under varying conditions.

These studies collectively showcase the evolving landscape of landslide mapping techniques using CNNs. Key advantages of these approaches include improved accuracy in complex terrains, rapid processing of large datasets, reduced false positives through advanced feature learning, and the potential for near-real-time mapping in post-disaster scenarios. While these

methodologies show great promise for enhancing our understanding of landslide dynamics and improving risk assessment, their practical application in emergency scenarios remains a challenge. The reliance on extensive manual mapping and pre-existing datasets can hinder their effectiveness during critical response efforts. As ongoing research focuses on developing CNN-based methods tailored for emergency situations, there is a growing need to balance academic advancements with the real-world applicability of these models, ensuring they can be effectively utilized in urgent disaster management contexts.

1.3.3 Early Warning Systems with Deep Learning

In the realm of landslide risk management, two main approaches leveraging machine learning techniques have emerged for practical application in emergency scenarios: local-scale and regional-scale predictions. Local-scale approaches focus on monitoring and predicting the behavior of specific, often slow-moving landslides, while regional-scale methods aim to assess landslide risk across larger areas. Despite the proliferation of studies in this field, it's important to note that many research efforts primarily focus on improving methodological aspects or performance metrics. Unfortunately, the translation of these academic advancements into operational emergency management systems remains limited. This gap between research and practice underscores the need for more emphasis on developing models that can be effectively integrated into real-world early warning systems and decision-making processes. Nevertheless, recent studies have made significant strides in both local and regional approaches, demonstrating the potential of machine learning in enhancing landslide risk assessment and prediction.

Local-Scale Predictions: Monitoring Slow-Moving Landslides

Landslide displacement prediction, particularly for slow-moving, is crucial for effective early warning systems. These landslides, marked by periods of stability and sudden acceleration, challenge traditional prediction methods due to the complex interactions between precipitation, reservoir water levels, and geological factors. Machine learning has become a valuable tool in this context, offering the ability to model the nonlinear relationships among these variables. However, traditional approaches often struggle with prediction instability and inaccuracy during critical periods of rapid deformation.

- **Zhou et al. (2018)** applied a hybrid model combining wavelet transform (WT) and Particle Swarm Optimization-Kernel Extreme Learning Machine (PSO-KELM) to predict the displacement of the Baishuihe landslide (Fig. 1.21a), located in the Three Gorges Reservoir Area (TGRA), China. This landslide was monitored over the period from 2003 to 2014, with detailed monthly displacement data. The study aimed to predict both the

trend and periodic displacement caused by internal geological factors and external triggers such as reservoir water levels and rainfall. The wavelet transform was used to denoise the data, and PSO was employed to optimize the KELM model. The model provided accurate forecasts with low error rates, making it suitable for use in landslide early warning systems.

- **Yang et al. (2019)** aimed to predict the displacement of slow-moving landslides in the Three Gorges Reservoir Area (TGRA) using a dynamic approach combining time series analysis and Long Short-Term Memory (LSTM) neural networks. The authors modeled landslide displacement by decomposing the time series data into a trend term and a periodic term (predicted using LSTM). The input data included monthly measurements of displacement and influencing factors like rainfall and reservoir water levels, focusing on two landslides, Baishuihe (Fig. 1.21bc) and Bazimen, over the period from 2004 to 2013. Importantly, their goal was to predict the displacement behavior, not landslide failure. The LSTM model provided more accurate predictions than traditional methods like SVM, particularly in capturing the dynamic movement of these step-wise landslides, highlighting its potential for improving early warning systems.

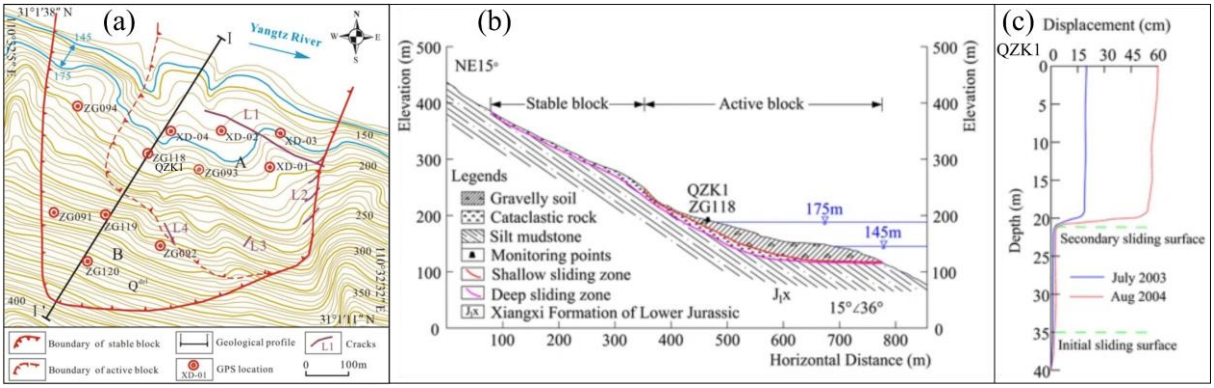


Figure 1.21. The topographical map in (**a**) provides an overview of the Baishuihe landslide area (modified from Zhou et al., 2018). (**b**) Schematic geological cross-section illustrates the internal structure of the Baishuihe landslide. (**c**) Displays the lateral displacement data versus depth, measured by inclinometer QZK1 (modified from Yang et al., 2019).

Both approaches have shown significant improvements over traditional methods, particularly in managing the nonlinearity and complexity of landslide behavior. Their accuracy during periods of rapid deformation is highly relevant for early warning systems, where timely predictions are critical for public safety. However, while these models hold great promise for

enhancing early warning capabilities, direct implementation in operational systems remains a challenge, requiring further research and development.

These advancements highlight the potential of combining sophisticated algorithms with a deep understanding of landslide dynamics. Yet, they also expose ongoing challenges, such as the need for extensive, high-quality monitoring data and the difficulty of generalizing models across varied geological settings. Future research may focus on reducing data dependency or integrating multiple data sources to provide a more comprehensive representation of landslide behavior.

Regional-Scale Predictions: Early Warning over Large Areas

At the regional scale, machine learning models are increasingly being employed to assess landslide risk across extensive areas, providing valuable tools for large-scale disaster prevention and planning (see Chapter 2).

- **Osanai et al. (2010)** introduced a nationwide early-warning system in Japan for predicting debris flows and slope failures using rainfall indices. The system, operational since 2007, employs a Radial Basis Function Network (RBFN) to set thresholds for disaster occurrences based on short-term (60-minute cumulative rainfall) and long-term (soil-water index) rainfall indices. The dataset used to develop the system includes non-occurrence rainfall data, allowing the model to predict landslide risks across a 5-km grid mesh covering Japan. The model's key feature is its ability to set critical lines (CL) separating high and low-risk areas, providing local governments with advanced warnings to issue evacuation orders. The system, which was operationally validated in 2009, demonstrated its effectiveness in reducing the impacts of landslide disasters through timely alerts broadcasted via TV, radio, and the internet. However, while the system has proven valuable in operational settings, the authors acknowledge that accurately forecasting rainfall, especially in mountainous regions, remains a limitation.

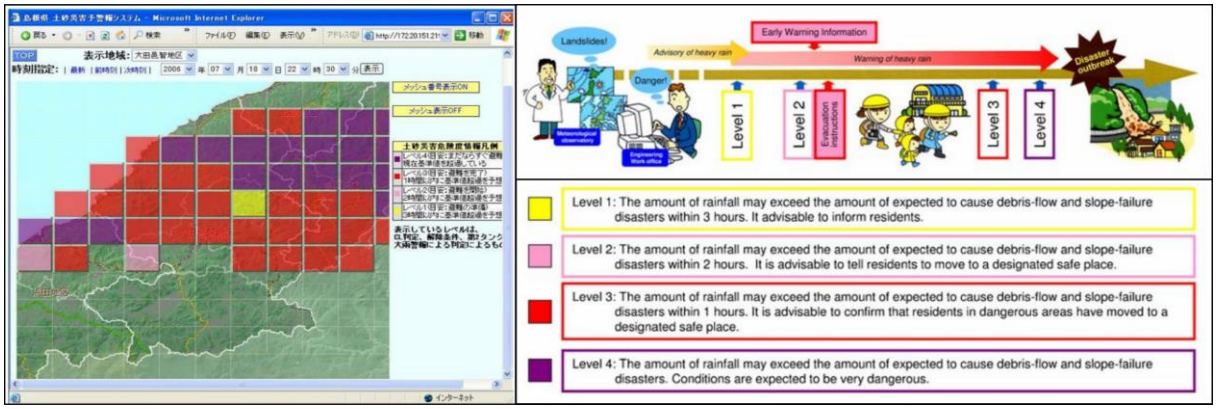


Figure 1.22. The format used by local governments to disseminate early warning information (left). The colors represent different danger levels, with corresponding recommended actions indicated below (Osanai et al., 2010).

- **Magrì et al. (2024)** developed a machine learning model to assess rainfall-induced landslide risk at a regional scale, focusing on the Liguria region in Italy. The study utilized a Polynomial Kernel Regularized Least Squares (KRLS) regression model to predict the daily occurrence of shallow landslides, based on critical hydrological factors such as soil moisture, 12-hour accumulated rainfall, and 3-hour rainfall peaks. The dataset, covering 2014 to 2019, included 2,191 recorded landslides across five Alert Zones (AZs), with data sourced from rainfall gauges and reports from the Liguria Civil Protection Department. The model was trained using only rainfall events that triggered landslides, focusing on critical conditions. The model's outputs were aligned with the operational early warning system used by the Liguria Civil Protection, offering a practical application for disaster risk management.

These larger-scale approaches demonstrate how machine learning can bridge the gap between scientific research and practical landslide risk management. These methods show promise for enhancing early warning systems and improving landslide risk mitigation across large areas, though continued refinement and validation across diverse geological contexts remain important areas for future work.

In conclusion, while both local and regional-scale approaches show significant promise in improving landslide prediction and risk assessment, their practical implementation in emergency scenarios remains an ongoing challenge. Future research should focus not only on improving model performance but also on developing frameworks that facilitate the integration of these advanced techniques into operational emergency management systems.

1.3.4 Emergency Response and Risk Assessment with CNN

These studies collectively showcase the evolving landscape of landslide mapping techniques using CNNs. Key advantages of these approaches include improved accuracy in complex terrains, rapid processing of large datasets, reduced false positives through advanced feature learning, and the potential for near-real-time mapping in post-disaster scenarios. However, challenges remain, such as the need for large, diverse training datasets and the computational resources required for processing high-resolution imagery. Additionally, some studies rely on pre-existing datasets and a 70-30 split between training and testing data, which, while important academically, poses practical challenges in emergency situations. In such cases, the model

would require pre-mapping 70% of the affected area to accurately map the remaining 30%, which is impractical when immediate mapping is essential for supporting rapid response efforts.

In light of these challenges, our focus has been on developing CNN-based methodologies specifically designed for emergency scenarios, where rapid and efficient mapping is crucial for disaster response and mitigation efforts. These advancements aim to enhance the practical applicability of landslide detection models in real-world situations, ensuring that they can be effectively utilized during critical times. To address these challenges in emergency scenarios, various strategies can be employed based on different approaches:

- *Create a Local Training Dataset (Supervised Learning)*: This approach (explained in section 1.2.3 and used in Chapter 3 and Chapter 5) involves manually mapping a portion of the affected area to create a training dataset. Meena et al. (2022) focused on automated landslide detection in the Rasuwa District of Nepal, which was affected by the 2015 Gorkha earthquake, triggering a total of 239 manually mapped landslide polygons. They utilized RapidEye satellite imagery with a spatial resolution of 5 meters, captured on November 4, 2016, and divided the dataset into training (23% or 55 polygons) and testing (77%) zones. Within the training polygons, 117 sampling points were selected from landslide areas and 57 from non-landslide regions, resulting in a total of 174 sampling points. The U-Net model achieved an F1-score of 71.12%, demonstrating its effectiveness in landslide segmentation while addressing the challenges of small training datasets.
- *Utilize Pre-trained Models*: This approach leverages models pre-trained on large, diverse datasets, adapting them for specific tasks like landslide detection. Prakash and Manconi (2021) applied this method using Sentinel-2 imagery to rapidly map landslides triggered by Storm Alex in October 2020. Their study utilized a generalized Convolutional Neural Network (CNN) trained on multiple landslide inventories, enabling quick mapping of 1,249 landslides. The advantage of this pre-trained model lies in its exposure to a vast amount of data, enhancing its ability to generalize across various scenarios. However, a significant limitation is that many pre-trained models are primarily trained on RGB imagery, which may restrict their effectiveness in detecting landslides in diverse environments. Additionally, while landslide-specific datasets are not very common, there has been a recent increase in their availability. Resources like Papers with Code show that more landslide datasets are emerging, yet the scarcity of comprehensive pre-trained models remains a challenge for effective application in this domain.

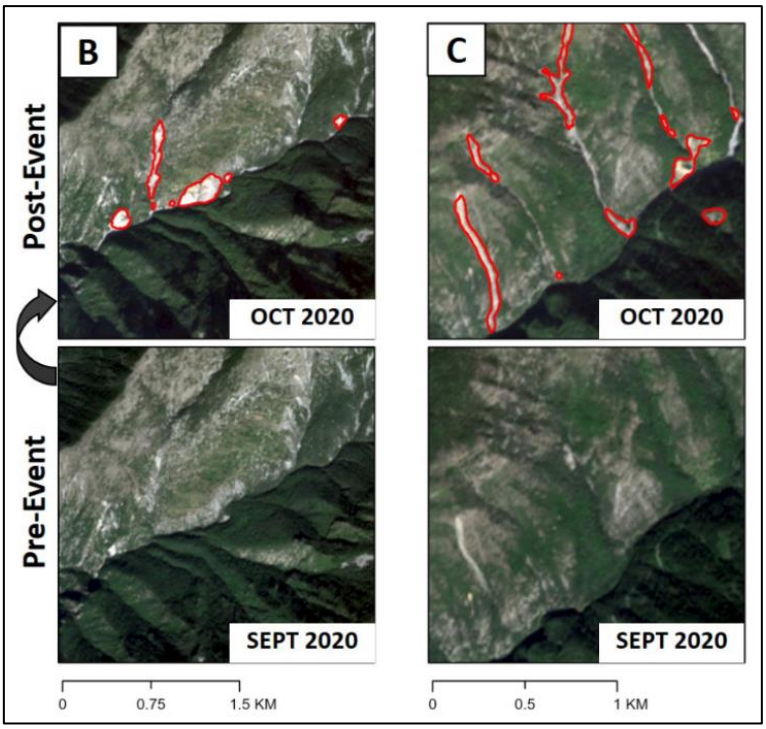


Figure 1.25. Landslide mapping results using generalized CNN. (B) and (C) display changes observed in Sentinel-2 images before and after the triggering event, with landslides identified by the CNN marked as red polygons in the post-event images. These examples highlight the effectiveness of CNN in detecting landslides, with (C) showing a typical detection scenario (Prakash and Manconi, 2021, modified).

- *Semi-supervised Learning:* Combining small amounts of labeled data with large amounts of unlabeled data, semi-supervised learning can enhance model performance when labeled training data is limited. This approach can be particularly effective in integrating newly acquired post-disaster images with existing datasets to improve mapping accuracy (see section 1.2.3).
- *Unsupervised Learning:* Shahabi et al. (2022) explored an unsupervised deep learning approach for rapid landslide detection using Sentinel-2 data. Their method combined a Convolutional Auto-encoder (CAE) with Mini Batch K-means clustering, achieving a mean Intersection over Union (mIOU) of 70% without requiring labeled training data. The advantage of this approach lies in its ability to utilize vast amounts of unlabeled data, making it particularly valuable in scenarios where obtaining labeled data is challenging or time-consuming (see section 1.2.3). However, the model's accuracy may vary based on the quality of the input data, and the absence of labeled datasets for validation can limit the assessment of its performance. This study demonstrates the potential of unsupervised methods for rapid landslide mapping in emergency situations, highlighting the need for further exploration in this area.

- *Few-shot and Zero-shot Learning:* These approaches aim to generalize from few examples or identify new classes without prior exposure. (section 1.2.7). While their application in landslide mapping is still limited, these methods have shown promising results in various fields of semantic segmentation. Notably, the study by Xi et al. (2024) presents the SAM-CFFNet model, which employs a cross-feature fusion network for the intelligent identification of landslides. This approach demonstrates the potential of zero-shot learning by effectively identifying landslides without requiring extensive labeled datasets. The model's ability to generalize from minimal examples offers a significant advantage in rapidly adapting to new types of landslides or different geographical areas. However, challenges remain in ensuring the model's accuracy and robustness in unfamiliar contexts where it has not been trained.

By developing and refining these methodologies, future research can ensure that CNN-based landslide detection models are not only academically robust but also practically applicable in real-world emergency scenarios, ultimately enhancing disaster response and mitigation efforts.

1.4 Research questions and outline

In the previous sections, the application of machine learning (ML) in landslide prediction and emergency management has been discussed as a promising yet challenging approach. Despite advancements in ML techniques and the increasing availability of high-resolution data, effectively applying these methods in real-world scenarios—where geological, meteorological, and topographical conditions are highly variable—remains a significant challenge. Key issues include identifying reliable rainfall thresholds for landslide initiation, integrating ML models into operational early warning systems, and utilizing automated mapping techniques in time-sensitive emergency situations. These challenges form the core of the research questions addressed in this PhD project.

The Emilia-Romagna region serves as a critical context for this research, known for its complex geological structure and frequent landslides, particularly exacerbated by extreme weather events. The significant landslide event in May 2023 underscored the urgent need for efficient landslide prediction and rapid response mapping, especially in light of climate change and its influence on precipitation patterns. This research project aims to explore how machine learning can improve our ability to predict and manage landslide risks in such complex environments, thus addressing the previously mentioned challenges. The following chapters will delve into the specific research questions, using case studies from the Emilia-Romagna region to illustrate the application of these methodologies in real-world scenarios.

Three main research questions have guided this project:

Question 1: *How can we effectively apply ML methods in complex geological conditions?*

This question will be addressed in Chapter 2, which focuses on the application of ML models to define rainfall thresholds for landslides in the Emilia-Romagna region. This chapter compares traditional statistical methods with ML approaches, assessing their effectiveness in improving landslide prediction under challenging geological conditions.

Question 2: *Is it truly effective to use ML methods for mapping during emergency situations?*

This will be explored in Chapter 3 and Chapter 5, where the application of deep learning models (e.g., U-Net and SegFormer) for rapid landslide detection during the 2023 emergency is discussed. These chapters examine the potential and limitations of automated mapping methods in real-time disaster response.

Question 3: *How can ML models be integrated into operational early warning and emergency response systems?*

This question is addressed by developing an integrated interface for the Emilia-Romagna Region's early warning system, which combines machine learning (ML) models with existing frameworks (Chapter 6). By augmenting conventional methods with ML insights, we aim to enhance decision-making processes and improve the overall effectiveness of emergency response strategies. Chapter 6 further concludes by synthesizing these findings and outlining future research directions.

1.5 References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160.

Angeli, M.G., Pasuto, A., & Silvano, S. (2000). A critical review of landslide monitoring experiences. *Engineering Geology*, 55(3), 133-147. https://doi.org/10.1016/S00137952(99)00122-2

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.1409.0473

Ball, N.M., & Brunner, R.J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07), 1049-1106. https://doi.org/10.1142/S0218271810017160

Baum, R.L., & Godt, J.W. (2010). Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides*, 7(3), 259-272. https://doi.org/10.1007/s10346-009-0177-0

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Computational Biology*, 4(10), e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166. https://doi.org/10.1109/72.279181

Berti, M., Martina, M.L.V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface*, 117(F4). https://doi.org/10.1029/2012JF002367

Bhowmick, A., & Hazarika, S.M. (2018). Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends. https://doi.org/10.48550/arXiv.1606.01042

Bishop, C.M. (2006). Pattern Recognition and Machine Learning. *Springer*.

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). Classification and regression trees. *CRC press*.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brown, T.B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Brynjolfsson, E. and Mcafee, A. (2017). The Business of Artificial Intelligence. *Harvard Business Review*, 7, 3-11. https://starlab-alliance.com/wp-content/The-Business-of-Artificial-Intelligence.pdf (accessed 14 Aug 2024)

Chapelle, O., Scholkopf, B., & Zien, A. (2006). Semi-Supervised Learning. *MIT Press.*

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, L., et al. (2021). Decision transformer: Reinforcement learning via sequence modeling. *Advances in Neural Information Processing Systems*, 34, 15084-15097. https://doi.org/10.48550/arXiv.2106.01345

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. https://doi.org/10.48550/arXiv.1406.1078

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018

Cox D. R., The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 20, Issue 2, July 1958, Pages 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Crosta, G.B., & Frattini, P. (2003). Distributed modelling of shallow landslides triggered by intense rainfall. *Natural Hazards and Earth System Sciences*, 3(1/2), 81-93. https://doi.org/10.5194/nhess-3-81-2003

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303-314. https://doi.org/10.1007/BF02551274

Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of*

the Association for Computational Linguistics: Human Language Technologies, 1, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Ding, A., Zhang, Q., Zhou, X., & Dai, B. (2016). Automatic recognition of landslides based on CNN and texture change detection. *In Proceedings of the Chinese Association of Automation (YAC), Youth Academic Annual Conference* (pp. 444-448). IEEE. https://doi.org/10.1109/YAC.2016.7804935

Dong, L., Xu, S. and Xu, B., (2018). Speech-Transformer: A No-Recurrence Sequence-to-Sequence Model for Speech Recognition. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, 2018, pp. 5884-5888, https://doi.org/10.1109/ICASSP.2018.8462506

Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929

Draper, N.R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). *John Wiley & Sons*.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211. https://doi.org/10.1207/s15516709cog1402_1

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *In Proceedings of the 34th International Conference on Machine Learning* (pp. 1126-1135).

Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé III, H., & Crawford, K. (2018). *Datasheets for datasets*. https://doi.org/10.48550/arXiv.1803.09010

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S., 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

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/s10346-021-01809-z

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680. https://doi.org/10.48550/arXiv.1406.2661

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press*, Cambridge, MA. Available online: https://www.deeplearningbook.org/ (Accessed on 26 June 2007).

Graves, A. (2012). Supervised Sequence Labelling with Recurrent Neural Networks. *Springer*. https://doi.org/10.1007/978-3-642-24797-2

Gunning, D., & Aha, D.W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3), 239-267. https://doi.org/10.1007/s00703-007-0262-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2008). The rainfall intensity-duration control of shallow landslides and debris flows: an update. *Landslides*, 5(1), 3-17. https://doi.org/10.1007/s10346-007-0112-1

Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M., & Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Science Reviews*, 200, 102973. https://doi.org/10.1016/j.earscirev.2019.102973

Guzzetti, F., Melillo, M., Calvello, M. et al. (2024). Independent demonstration of a deep-learning system for rainfall-induced landslide forecasting in Italy. *Landslides*. https://doi.org/10.1007/s10346-024-02294-w

Hartigan, J.A., & Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100-108. https://doi.org/10.2307/2346830

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Hinton, G.E., Osindero, S., & Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527

Ho, T.K. (1995). Random Decision Forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 278-282. https://doi.org/10.1109/ICDAR.1995.598994

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257. https://doi.org/10.1016/0893-6080(91)90009-T

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). Applied Logistic Regression (3rd ed.). *Wiley*.

Iverson, R.M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, 36(7), 1897-1910. https://doi.org/10.1029/2000WR900090

Jain, A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Jordan, M. I., & Bishop, C. M. (1996). Neural networks. In A. Tucker (Ed.), *CRC Handbook of Computer Science CRC Press.*

Jordan, M.I., & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Karthikeyan, A., Priyakumar, U.D. Artificial intelligence: machine learning for chemical sciences. *J Chem Sci* 134, 2 (2022). https://doi.org/10.1007/s12039-021-01995-2

Kirschbaum, D., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, 6(3), 505-523. 0 https://doi.org/10.1002/2017EF000715

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. https://doi.org/10.1145/3065386

Kutner, M.H., Nachtsheim, C.J., & Neter, J. (2004). Applied Linear Regression Models. *McGraw-Hill Irwin*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Lloyd, S.P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489

Loh, W.Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14-23. https://doi.org/10.1002/widm.8

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765- 4774.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

Magrì, S., Solimano, M., Delogu, F. et al. (2024). Modelling rainfall-induced landslides at a regional scale, a machine learning based approach. *Landslides*, 21, 573-582. https://doi.org/10.1007/s10346-023-02173-w

Meena, S.R., Ghorbanzadeh, O., van Westen, C.J., Nachappa, T.G., Blaschke, T., Singh, R.P., & Sarkar, R. (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., van Westen, C., Bhuyan, K., Singh, R.P., & Catani, F. (2022). Landslide detection in the Himalayas using machine learning algorithms and U-Net. *Landslides*, 19(5), 1209-1225. 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

Menard, S. (2002). Applied Logistic Regression Analysis (2nd ed.). *Sage*.

Mikhaylov SJ, Esteve M, Campion A. (2018). Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. *Phil. Trans. R. Soc. A* 376: 20170357. http://dx.doi.org/10.1098/rsta.2017.0357

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech*, 2, 1045-1048. https://doi.org/10.21437/Interspeech.2010-343

Mitchell, T.M. (1997). Machine Learning. *McGraw-Hill*.

Mondini, A.C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, 14(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective. *MIT Press*.

Nair, V., & Hinton, G.E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807-814.

Nava, L., Bhuyan, K., Meena, S.R., Monserrat, O., & Catani, F. (2022). Rapid mapping of landslides on SAR data by attention U-Net. *Remote Sensing*, 14(6), 1449. https://doi.org/10.3390/rs14061449

Nocentini, N., Rosi, A., Piciullo, L. et al. (2024). Regional-scale spatiotemporal landslide probability assessment through machine learning and potential applications for operational warning systems: a case study in Kvam (Norway). *Landslides* 21, 2369–2387. https://doi.org/10.1007/s10346-024-02287-9

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1717-1724. https://doi.org/10.1109/CVPR.2014.222

Osanai, N., Shimizu, T., Kuramoto, K. et al. (2010). Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network. *Landslides* 7, 325–338. https://doi.org/10.1007/s10346-010-0229-5

Pan, S.J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Petley, D.N., Bulmer, M.H., & Murphy, W. (2005). Patterns of movement in rotational and t ranslational landslides. *Geology*, 33(2), 161-164. https://doi.org/10.1130/0091-7613

Piciullo, L., Calvello, M., & Cepeda, J.M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, 179, 228-247. https://doi.org/10.1016/j.earscirev.2018.02.013

Prakash N. and Manconi A. (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

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106. https://doi.org/10.1007/BF00116251

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint* arXiv:1511.06434. https://doi.org/10.48550/arXiv.1511.06434

Radford, A., et al. (2018). Improving language understanding by generative pre-training. *OpenAI Blog*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). https://doi.org/10.1109/CVPR.2016.91

Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386-408. https://doi.org/10.1037/h0042519

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. https://doi.org/10.1038/323533a0

Samek, W., Wiegand, T., & Müller, K.R. (2023). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Model. *Springer Nature*. https://doi.org/10.48550/arXiv.1708.08296

Samuel, A.L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229. https://doi.org/10.1147/rd.33.0210

Schölkopf, B., & Smola, A.J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. *MIT press*. https://doi.org/10.7551/mitpress/4175.001.0001

Seber, G.A.F., & Lee, A.J. (2012). Linear Regression Analysis. *John Wiley & Sons*.

Segoni, S., Piciullo, L., & Gariano, S.L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15(8), 1483-1501. https://doi.org/10.1007/s10346-018-0966-4

Shahabi H. et al., (2022). Rapid Mapping of Landslides from Sentinel-2 Data Using Unsupervised Deep Learning. *IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, Istanbul, Turkey, 2022, pp. 17-20, https://doi.org/10.1109/M2GARSS52314.2022.9840273

Sidle, R.C., & Ochiai, H. (2006). Landslides: processes, prediction, and land use. *American Geophysical Union.* http://dx.doi.org/10.1029/WM018

Silver, D., Huang, A., Maddison, C. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. https://doi.org/10.1038/nature16961

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. https://doi.org/10.48550/arXiv.1312.6034

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *In Advances in Neural Information Processing Systems* (pp. 4077-4087). https://doi.org/10.48550/arXiv.1703.05175

Stanley T.A., Kirschbaum D.B., Benz G., Emberson R.A., Amatya P.M., Medwedeff W. and Clark M.K. (2021). Data-Driven Landslide Nowcasting at the Global Scale. *Front. Earth Sci.* 9:640043. https://doi.org/10.3389/feart.2021.673644

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. https://doi.org/10.48550/arXiv.1906.02243

Sun, T.Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Sutskever, I., Vinyals, O., & Le, Q.V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 3104-3112.

Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction (2nd ed.). *MIT Press.*

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. *In Advances in Neural Information Processing Systems* (pp. 3630-3638).

Werbos, P.J. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. *Harvard University*.

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

Xi, L.; Yu, J.; Ge, D.; Pang, Y.; Zhou, P.; Hou, C.; Li, Y.; Chen, Y.; Dong, Y. (2024). SAM-CFFNet: SAM-Based Cross-Feature Fusion Network for Intelligent Identification of Landslides. *Remote Sens*, 16, 2334. https://doi.org/10.3390/rs16132334

Xian, Y., Lampert, C.H., Schiele, B., & Akata, Z. (2018). Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9), 2251-2265. https://doi.org/10.48550/arXiv.1707.00600

Xie, E., et al. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34, 12077-12090. https://doi.org/10.48550/arXiv.2105.15203

Yang, B., Yin, K., Lacasse, S., & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides*, 16, 677-694. https://doi.org/10.1007/s10346-018-01127-x

Yu, B., Chen, F., & Xu, C. (2020). Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Computers & Geosciences*, 135, 104388. https://doi.org/10.1016/j.cageo.2019.104388

Yu, M., Yang, C., & Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences*, 8(5), 165. https://doi.org/10.3390/geosciences8050165

Zhao, Y., Meng, X., Qi, T., Li, Y., Chen, G., Yue, D., & Qing, F. (2022). AI-based rainfall prediction model for debris flows. *Engineering Geology*, 296, 106456. https://doi.org/10.1016/j.enggeo.2021.106456

Zhou, C., Yin, K., Cao, Y., Ahmed, B., Intrieri, E., & Catani, F. (2018). Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. *Landslides*, 15(11), 2211-2225. https://doi.org/10.1007/s10346-018-1022-0

Zhu, X., & Goldberg, A. B. (2009). Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 3(1), 1-130. DOI: https://doi.org/10.2200/S00196ED1V01Y200906AIM006

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *Lecture Notes in Computer Science*, *8689*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Chapter 2

Comparative analysis of conventional and machine learning techniques for rainfall threshold evaluation under complex geological conditions

2.1 Preface

This chapter focuses on a crucial aspect of hydrogeological risk management: the definition of reliable rainfall thresholds for predicting rain-induced landslides in complex geological contexts. The presented work represents a significant step in my doctoral research, which aims to improve early warning systems for landslide phenomena in the Emilia-Romagna region.

The research centered on comparing conventional methods with machine learning techniques for evaluating rainfall thresholds, using the Emilia-Romagna region as a case study. This area is known for its complex geological structure and the prevalence of weak rocks prone to large-scale landslides. Despite the availability of an extensive historical dataset on landslides and precipitation, defining reliable rainfall thresholds in this context proved to be a significant challenge, requiring a multidisciplinary approach and the application of advanced analytical techniques.

This chapter presents the results of this comparative analysis, offering insights not only into the relative effectiveness of different methods but also into future directions for improving rain-induced landslide prediction. The conclusions reached have implications for the development of more effective warning systems and for understanding the key factors influencing landslide triggering in complex geological contexts (Chapter 6).

The chapter is based on the article published in *Landslides* journal:

Dal Seno N.[1], Evangelista D.[2], Piccolomini E.[2], Berti M.[1], 2024. Comparative analysis of conventional and machine learning techniques for rainfall threshold evaluation under complex geological conditions, https://doi.org/10.1007/s10346-024-02336-3.

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

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

2.2 Abstract

This research focuses on the essential task of defining rainfall thresholds in regions with complex geological features, specifically at a regional scale. It examines a variety of methodologies, from traditional empirical-statistical methods to cutting-edge machine learning (ML) techniques, for establishing these thresholds. The Emilia-Romagna region in Italy, known for its intricate geological structure and prevalence of weak rocks that often lead to large and deep-seated landslides, serves as the study area. The region's complex interplay between rainfall and landslide incidences poses a significant challenge in accurately determining rainfall thresholds. The effectiveness of ML methods is compared against conventional empirical-statistical approaches, evaluating factors such as prediction accuracy, model complexity, and the interpretability of results for use by regional landslide warning system operators. The findings indicate that machine learning techniques have an edge over traditional methods, yielding higher performance scores and fewer false positives. Nevertheless, these advancements are modest when considering the increased complexity of ML methods and the incorporation of additional rainfall parameters. This underlines the continued need for improvements in data quality and volume. The study stresses the importance of enhancing data collection and analysis techniques, especially in an era where advanced AI tools are increasingly available, to improve the accuracy of predicting rainfall thresholds for effective landslide warning systems.

2.3 Introduction

The definition of rainfall thresholds is critical for developing effective landslide warning systems, both regionally and locally (De Vita et al. 1998; Aleotti et al. 1999; Wieczorek and Glade 2007; Guzzetti et al. 2007a, 2008). Rainfall is the primary trigger of landslides worldwide (Caine 1980) and identifying the critical rainfall conditions that trigger landslides is crucial for predicting their occurrence (Terlien, 1998; Calvello & Piciullo, 2016). Various methods exist for setting these thresholds, including empirical-statistical and physically-based approaches (Segoni et al., 2018; Guzzetti et al., 2007b). Empirical-statistical methods, which analyze historical landslide and rainfall data to establish a relationship, are widely used due to their simplicity and robustness (Berti et al., 2012; Aleotti 2004; Glade et al., 2000). They account for complex interactions affecting slope stability but are often specific to certain locations and may not hold up under changing climatic conditions (Crosta & Frattini, 2022; Gariano & Guzzetti, 2016).

Machine learning (ML) techniques have brought significant advancements to this field. These algorithms learn from existing data to identify patterns and predict future events, offering adaptability and versatility in analyzing various variables (Alpaydin, 2010; Bishop, 2006). Recent studies have demonstrated the effectiveness of machine learning in improving landslide

warning systems by integrating various data types, including soil moisture and rainfall, to predict hydrologic responses and landslide occurrences (Orland et al., 2020, Liu et al. 2021, Mondini et al. 2022). These techniques represent a promising direction for enhancing the accuracy and applicability of rainfall thresholds in landslide prediction.

This study assesses the use of machine learning techniques for determining rainfall thresholds in Emilia-Romagna, Northern Italy, a region characterized by complex geology and weak rocks that lead to frequent, large landslides. The research focuses on evaluating the effectiveness of machine learning approaches in comparison with traditional empirical-statistical methods. Key aspects of this evaluation include prediction accuracy, model complexity, and the ease of interpreting and communicating findings to regional warning system staff, considering the intricate relationship between rainfall and landslides in the area.

2.4 Study area

The Emilia-Romagna region in Italy (Fig. 2.1) is highly prone to landslides due to several factors. The region's rocks are mostly weak and susceptible to weathering, and there is ongoing geological uplift, constantly reshaping the landscape. The lengthy rainy and snowy seasons further contribute to slope instability. Over 32,000 landslides have been identified in the area (Bertolini et al., 2005), with earthflows being the most prevalent type, especially in regions with clay rocks (Ligurian-Subligurian Domain, Fig. 2.1). These earthflows and slides typically move slowly, but can accelerate dramatically during heavy rain or snowmelt, sometimes reaching high speeds. Rockslides, another common type, are frequent in the flysch units (Tuscan-Romagna Domain, Fig. 2.1) and can form large, dangerous slides, especially when strata dip out of the slope. While less common than earthflows, rockslides pose a greater risk due to their higher speeds. Rockfalls occur in competent sandstone and limestone rocks (Epiligurian Domain, Fig. 2.1) but represent less than 1% of landslides in the region. Figure 2.2 provides illustrative examples of these types of landslides.

The Civil Protection Agency of Emilia-Romagna is responsible for managing the region's hydrogeological risk alert system. Since 1995, the agency has established and continuously improved the alert system, which comprises of a prediction and monitoring phase. The prediction phase relies on meteorological forecasts for the subsequent 72 hours and an evaluation of their corresponding effects. A team of specialized experts performs this crucial step, considering various rainfall threshold models and the current territory conditions. The technical group's assessment is reported in a daily Alert Bulletin that identifies three possible alert levels (ordinary, moderate, elevate) based on the expected number and extent of landslides. Such indications enable the implementation of preparation plans at the municipality level. The forecasts cover eight alert areas (Fig. 2.1), which are territorial units characterized by homogenous meteo-climatic response and comparable hydrographic and orographic features.

Alert areas are considerable in size (from 660 to 1900 km^2) and exhibit significant geological variability within their boundaries, with different types of landslide phenomena. Nonetheless, the regional alert system aims to prepare municipalities for potential landslides on a large scale, rather than providing forecast at the local level.

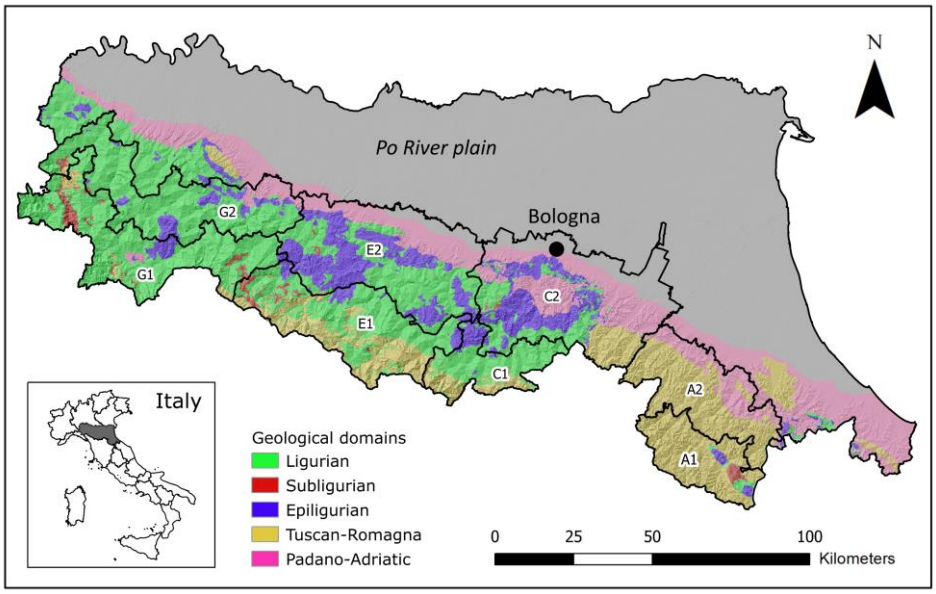


Figure 2.1. Geographic overview of the Emilia-Romagna region (Italy) segmented into alert zones. Illustrative geological map of the region.

Figure 1.2. Examples of landslide types in the Emilia-Romagna Region. (**a**) Earthflows and (**b**) earth slides, commonly found in the clay rocks of the Ligurian Domain. (**c**) A typical rockslide observed in the flysch rocks of the Tuscan-Romagna Domain. (**d**) Rockfall characteristic of the sandstone rocks in the Epiligurian Domain.

2.5 Methods

2.5.1 Data

The analysis in this study was based on data from two primary sources: a historical catalogue of landslide events and precipitation records. A similar dataset was previously used by Berti et al. (2012) to determine probabilistic rainfall thresholds for Emilia-Romagna using a Bayesian approach. This current study differs in three key aspects: i) it includes updated data on landslides from 2009 to 2022; ii) it adopts a more representative approach for averaging rainfall over alert areas; iii) it evaluates various traditional and machine learning algorithms for defining rainfall thresholds.

Landslide catalogue

The Emilia-Romagna Geological Survey has a register of past landslides occurred in the regional territory (https://ambiente.regione.emilia-romagna.it/en/geologia/cartography/). The register is constantly updated and includes data from national archives, local press, and technical reports. Although the catalogue is not entirely comprehensive, since minor events that occurred in remote areas may have gone unnoticed, it can be regarded as statistically complete from approximately 1951 (Rossi et al., 2010). The catalogue stores the location, date of occurrence, characteristics, and consequences of each landslide with varying degrees of accuracy. The date occurrence is the most important parameter for our analysis since each landslide must be linked to a specific triggering precipitation event. To ensure accuracy, only landslides with a certain date of occurrence (daily accuracy) or with a maximum uncertainty of 1-7 days were considered.

Rainfall data

The Emilia-Romagna region features a rainfall network with over 200 tipping-bucket rain gauges spread uniformly across the area. Up to 2001, rainfall data was collected manually on a daily basis, but since then, automatic gauges have been used to record data every 30 minutes. In the civil protection system, rainfall measurements from an entire alert area (as shown in Fig. 2.1) are spatially interpolated and averaged to provide a single value for the whole area, due to the variability of precipitation over short distances. The averaging of daily rainfall data also matches with weather forecasts, which predict precipitation levels for each alert zone over 24, 48, and 72 hours. The alert zones vary in size, from 660 km² (Zone A1) to 1900 km² (Zone C2), with each rain gauge monitoring an area of approximately 60 km².

Manual identification of triggering rainfall events

Assigning a specific rainfall event to each historical landslide proved to be a challenging task. The complexity of rainfall, which can consist of multiple episodes with varying intensities and durations, along with the impact of snowmelt not included in rainfall data, contributed significantly to this difficulty. The uncertainty surrounding the date of the landslide occurrence further complicated the procedure, making it impossible to utilize an automatic algorithm for identifying the triggering rainfall event. Therefore, a manual identification process was employed, requiring careful analysis of each single case. This process was complex, time-consuming, and involved the continuous re-analysis of more than 4000 historical records.

To address the subjectivity inherent in expert judgment, each rainfall event was classified based on its uncertainty level (refer to Fig. 2.3 for examples):

- Type 1) Low uncertainty. The triggering rainfall can be identified with a minimal degree of subjectivity as there is a distinct rainfall episode that occurred during or, at most, 1-3 days prior to the landslide, with a clearly defined beginning and end (Fig. 2.3a-c).
- Type 2) Moderate uncertainty. At the time of the landslide there is a heavy rainfall episode, but determining the exact beginning of the event is somewhat subjective due to the presence of multiple rainfall pulses with comparable duration and intensity (Fig. 2.3d-f).
- Type 3) High uncertainty. Determining the triggering rainfall with sufficient accuracy is not possible due to several reasons, such as: a) the landslide occurred at the beginning of a rainfall event (Fig. 2.3g); b) the landslide occurred during a dry period, or after a light rainfall with minimal duration and intensity (Fig. 2.3h); c) the landslide occurred after or during an extended period characterized by multiple rainfall events with similar duration and intensity (Fig. 2.3i). These records may indicate landslides with an incorrect date of occurrence, landslides triggered by snowmelt or rain-on-snow events, or situations where the relationship between rainfall and the landslide is particularly intricate because of the effect of antecedent rainfall.

For analytical purposes, Type 1 and Type 2 events are considered appropriate, while Type 3 events are excluded due to the high uncertainty about the duration and intensity of the triggering rainfall.

Regarding rockfalls, they were not explicitly excluded from our analysis; however, any events not clearly triggered by rainfall were categorized as Type 3 'High uncertainty' during the manual identification process and were consequently excluded from further analysis.

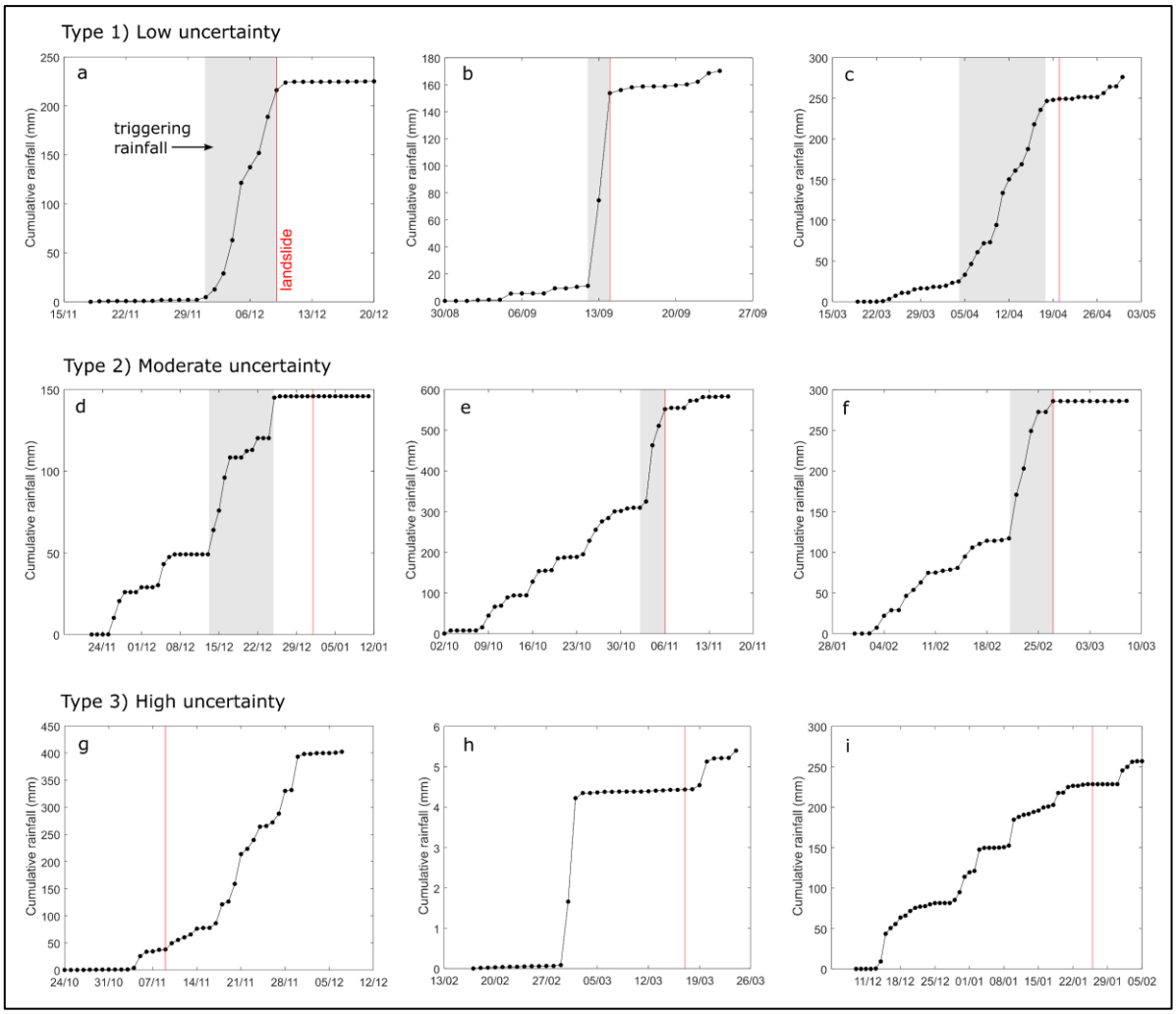


Figure 2.3. Classification of Rainfall Events by Uncertainty in Expert Assessment. (**a-c**) Type 1 - Low Uncertainty: The triggering rainfall is clearly identified with minimal subjectivity, evident from a distinct rainfall episode occurring 1-3 days before the landslide. (**d-f**) Type 2 - Moderate Uncertainty: Heavy rainfall coincides with the landslide, but discerning the precise start is uncertain due to multiple similar-duration and intensity rainfall pulses. (**g-i**) Type 3 - High Uncertainty: Pinpointing the triggering rainfall is challenging; the landslide could occur at the onset of a rainfall event (g), during a dry period (h), or after an extended period with several similar rainfall events (i).

Automated definition of rainfall events from precipitation data

To define rainfall thresholds via statistical techniques (discussed in the following section), it's essential to categorize both critical and non-critical rainfall events. Given the extensive span of the precipitation data, an automated process is necessary for isolating non-significant rainfall events within the dataset. A common approach in literature is to set a threshold of duration ($D_{\min}$) and precipitation amount ($P_{\min}$) which serves to determine the start and end of a rainfall

event (Corominas 2000, Saito et al. 2010). The values of D_{min} and P_{min} depend on the specific site conditions: short inter-event periods are typically used for shallow landslides and debris flows, that promptly react to rainfall, while longer periods are employed for deep-seated landslides in fine-grained materials (Wang et al. 2003). Regardless of the criteria used, it is essential to ensure consistency between the automated algorithm and the expert identification of critical rainfall. The algorithm, in fact, must be developed based on the same criteria and thresholds used by the experts to guarantee consistency between the datasets of critical and non-critical rainfall.

During the manual identification of the triggering rainfall, various combination of D_{min} and P_{min} were tested to determine which one was most effective in replicating expert judgment. Ultimately, the best combination was determined to be D_{min} =3 days and P_{min} =5 mm (D3P5 algorithm hereafter). This means that a rainfall event is defined as continuous period of precipitation that ends when the precipitation drops below 5 mm over 3 days. Berti et al. (2012) arrived at the same definition using a reduced version of the current dataset spanning the period 1939-2000. The D3P5 criterion was applied to the precipitation time series of the eight alert areas to extract all rainfall events in the period 1939-2022. Figure 2.4 shows an example of a rainfall time series (Alert Area A1, period February-September 1978) where rainfall events have been identified through the automatic recognition algorithm. The red polygon highlights a rainfall event that triggered landslides.

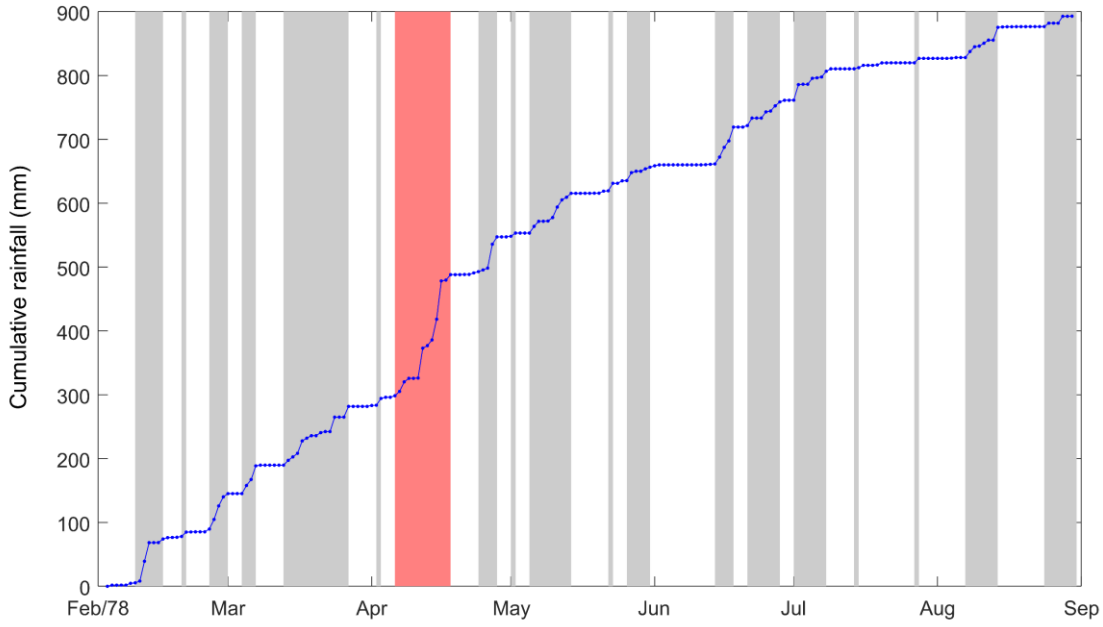


Figure 2.4. Demonstration of automatic rainfall event identification in the rainfall time series for Alert Area A1 (February-September 1978). The rainfall events within this period were identified using an automated recognition algorithm. An event that resulted in landslides is emphasized with a red polygon.

Dataset for analysis

The final dataset used for the analysis is organized into eight tables, each corresponding to a specific alert area. Every table row represents a distinct rainfall event, with various columns detailing key aspects of the event. These aspects include the duration of the event (in days), total precipitation (in millimeters), average daily rainfall intensity (mm/day), antecedent rainfall over preceding intervals of 7, 14, 30, and 60 days (in millimeters), and the day of the year (numbered 1-365, serving as an indicator of seasonal variation). Additionally, each row notes the number of landslides triggered in the alert area by the rainfall event, where a zero indicates a non-critical event that did not result in any landslides.

While all events are geolocated and theoretically geological data could be included in the analysis, the traditional threshold methods used in this study are typically based only on two parameters—rainfall duration and intensity/cumulative. Including geological data would necessitate subdividing the dataset into geologically homogeneous areas, which would significantly reduce its statistical robustness. For machine learning methods, although feasible, we chose not to include geological factors to maintain a direct comparison with traditional methods, ensuring a more straightforward evaluation framework.

The dataset's primary focus is to investigate the link between rainfall characteristics and the occurrence of landslides on a larger scale, rather than delving into the broader environmental factors that may influence landslide risk. Hence, the analysis omits spatial details like geological makeup, slope gradients, and land use patterns, which are generally pivotal for understanding landslide genesis on a more localized, slope-specific scale.

2.5.2 Algorithms

In our study, we compared four traditional rainfall threshold determination methods (BART, Frequentist, Percentile, LDA) with seven advanced Machine Learning (ML) techniques (NN, LR, MLDA, MQDA, RF, XGBoost, TPOT), detailed in Table 2.1. This distinction highlights the shift from traditional methods, which rely on basic parameters like rainfall duration and intensity, to more complex, multiparametric ML techniques, as listed in Table 2.1. While traditional methods are straightforward, ML techniques, evolving from statistical principles of traditional methods, offer greater analytical depth but are more complex to learn and apply.

2.5.2.1 Conventional methods

The traditional techniques considered in our analysis comprise four various Bayesian methods: BART, Frequentist, Percentile, and LDA.

BART

BART (Bayesian Analysis of Rainfall Thresholds) is a probabilistic method for defining rainfall thresholds (Berti et al., 2012, 2015). It calculates the probability of a landslide occurring in response to specific rainfall duration and intensity using a simplified application of Bayes' theorem as proposed by Berti et al. (2012):

$$P(A|B) \approx \frac{N(B \mid A)}{N_B} \quad [2.1]$$

where *N(B|A)* is the number of rainfall events of magnitude B that resulted in landslides, and N_B is the number of rainfall events of magnitude B. This approximation simplifies the computation, making it practical for implementation while maintaining the necessary accuracy for our analysis. Detailed derivations can be found in Berti et al. (2012).

Frequentist

The Frequentist method, developed by Brunetti et al. (2010) and based on Fisher's principles from 1936, analyzes the frequency of rainfall conditions historically leading to landslides, as shown in Figure 2.5c-d. It involves plotting triggering rainfalls on a logarithmic scale (Fig. 2.5c), using the least squares method to find a fitting line (depicted in blue), and applying Kernel Density Estimation, as per Silverman (1986), to calculate the probability density function of regression residuals with a Gaussian model. The final step is setting a rainfall threshold for a specific exceedance probability. This optimal exceedance probability was identified on the training dataset by testing all intercepts corresponding to the Gaussian distribution of rainfall events that resulted in landslides, while maintaining the slope of the best-fit line (blue line in Fig. 2.5c) which corresponds to 50% of the distribution. Based on the optimal linear residual (delta) score, the T20 line (red line in Fig. 2.5d), which represents 20% of the distribution, demonstrated superior performance on the training dataset and was subsequently employed during the testing phase.

Percentile

Guzzetti et al. (2007) employed a percentile method to determine the alert threshold for landslide occurrences in Central and Southern Europe. The percentile is a statistical technique used to determine the value below which a certain percentage of observations in a dataset fall (Moore and McCabe, 1993). To perform the analysis, the rainfall events that resulted in landslides are plotted in the ID chart and the duration axis is divided into bins (Fig. 2.5e). The data falling within each bin are then sorted in ascending order based on rainfall intensity, and the point corresponding to a given percentile (e.g., 5%) is calculated. The rainfall threshold is

finally obtained by linearly fitting the ID points of the considered percentile (Fig. 2.5f). For our study, after evaluating different percentiles, the 25th percentile was identified as the most effective threshold based on the optimal delta score.

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA), a dimensionality reduction and feature selection technique, aims to find a linear combination of features that effectively separates different classes (Fisher, 1936; Brown, 1998; Venables and Ripley, 2002). Originally proposed by Fisher, LDA uses sample mean vectors and class variances to identify the optimal direction for class separation. It achieves this by maximizing between-class variance (S_B) and minimizing within-class variance (S_W), a process visualized in Fig. 2.5g-h. The separation is optimized through solving a generalized eigenvalue problem, with the formula:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \qquad [2.2]$$

where W represents the projection vector. In our study, the best projection vector on W differentiates critical from non-critical rainfall, serving as the rainfall threshold in the ID plot.

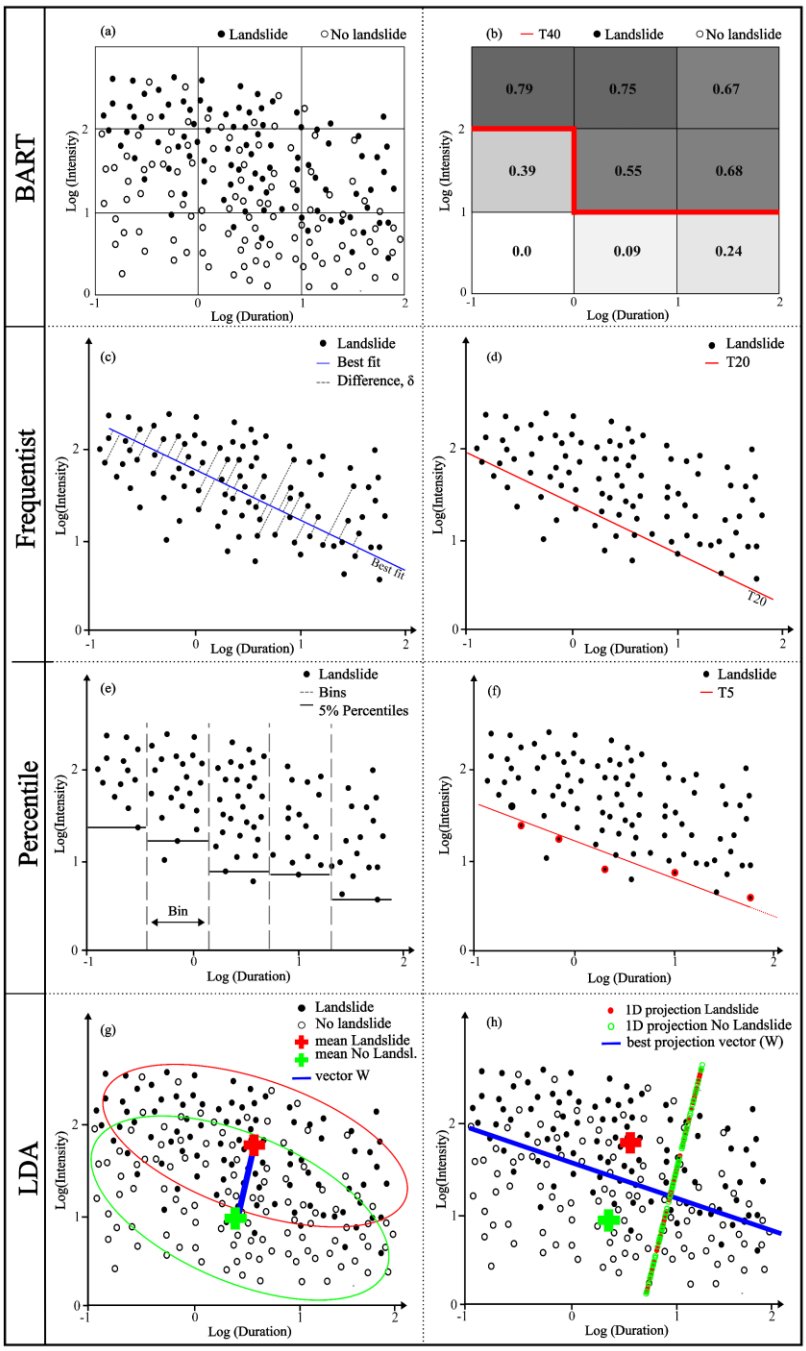


Figure 2.5. Conceptual examples of conventional methods for rainfall thresholds evaluation. BART: **a**) scatter plot of rainfall events that did and did not result in landslides; **b**) conditional probability plot of landslides for nine different combinations of intensity and duration, with the red line indicating the threshold with probability > 0.40. Frequentist: **c**) rainfall events that resulted in landslides plotted with the best-fit line (blue); **d**) rainfall threshold defined by shifting the regression line to the 20% exceedance probability (red line). Percentile: **e**) percentile plot of rainfall events showing the 5th percentile for each bin; **f**) rainfall threshold (T5) obtained from the best fit of the5th percentile points (outlined in red). LDA: **g**) mean vectors of critical and non-critical rainfall (red and green crosses) connected by the vector W (blue line); **h**) optimized projection vector W (blue line) for best class separation.

2.5.2.2 Machine Learning techniques

The four conventional methods were evaluated against seven algorithms from Machine Learning, using different numbers of input variables (Table 2.1).

Table 2.1. Overview of the methods used to define rainfall thresholds. Parameters: I=rainfall intensity (mm/day); D=rainfall duration (days); dY=day of the year (1 t o365); C=total precipitation (mm); A7, A14, A30, A60= antecedent rainfall in the previous 7, 14, 30 and 60 days (mm); Ac=Alert area.

Method	Parameters	Type of results	Reference
BART	I, D	Bayes probability	(Berti et al. 2012)
Frequentist	I, D	Exceedance probability	(Brunetti et al., 2010)
Percentile	I, D	Empirical distribution	(Guzzetti et al., 2007)
LDA	I, D	Binary classification	(Hastie et al., 2009)
NN	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Huang and Xiang 2018)
LR	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Frattini et al., 2009)
MLDA	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Ramos-Cañón et al., 2016)
MQDA	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Hastie et al., 2009)
RF	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Hu et al., 2021)
XGBoost	I, D, dY, C, A7, A14, A30, A60, Ac	Multiclass classification	(Zhao et al. 2022)
TPOT	I, D, dY, C, A7, A14, A30, A60, Ac	Optimized pipeline	(Olson et al. 2016)

Neural Networks (NNs)

Neural Networks (NN), machine learning methods inspired by the human brain, are adept at solving complex problems (Goodfellow et al. 2016; Schmidhuber 2014). They consist of layers of interconnected neurons, learning through weight adjustments for each input (Lecun et al., 2015). In the context of landslide prediction from rainfall data, NNs assign weights to factors such as duration and intensity, using them in a prediction process that involves normalization, weight combination, and activation functions like sigmoid, to output a probability indicating landslide likelihood (Fig 2.6a).

For our analysis, we implemented three NN architectures using the Keras Python library. The ‘shallow neural network’ had four layers with 32, 16, 8, and 1 neuron; the ‘medium neural

network' comprised five layers with 256, 64, 32, 8, and 1 neuron; and the 'large neural network' also had five layers but with 512, 128, 64, 8, and 1 neuron. All used `ReLU` activation for hidden layers and `sigmoid` for the output layer. They were trained with the Adam optimizer (learning rate 1e-4) and binary cross-entropy loss, considering class weights of 1 and 20 to balance data distribution between landslide and non-landslide classes.

Logistic Regression (LR)

Logistic Regression (LR) (Cox 1958; Hastie et al., 2009), is a generalized linear model using the sigmoid function to predict event probabilities. Like neural networks, LR assigns weights to input features during training and calculates the probability of a binary outcome by applying the sigmoid function to the sum of weighted inputs. LR aims to optimally set weights to minimize prediction errors, following the formula:

$$p(y = 1|x) = \frac{1}{(1 + e^{(-w^T x)})} \quad [2.3]$$

where $p(y=1|x)$ is the probability of the outcome being 1 given input x, with w as the weight vector.

In our study, LR was implemented for binary classification using Scikit-learn's `LogisticRegression` module with '`l2`' Ridge regularization. We chose '`lbfgs`' as solver to handle the optimization. To address class imbalance, we set `class_weight` to {0:1, 1:18.7}, reflecting the prevalence of class 0 over class 1. The LR model's conceptual approach (Fig. 2.6b) was effectively employed in our analysis.

Multiparametric Linear Discriminant Analysis (MLDA)

Multiparametric Linear Discriminant Analysis (MLDA) extends the concept of LDA (Fisher, 1936; Venables and Ripley, 2002). Unlike LDA, which considers only two predictor variables, MLDA accommodates multiple predictor variables, making it suitable for complex tasks like predicting rainfall-induced landslides.

In our study, we implemented MLDA using Scikit-learn's `LinearDiscriminantAnalysis` module with the '`lsqr`' solver, `priors` set to [0.4, 0.6], and a `shrinkage` parameter of 0.4 for regularization.

Multiparametric Quadratic Discriminant Analysis (MQDA)

Multiparametric Quadratic Discriminant Analysis (MQDA) extends MLDA for curved decision boundaries (Hastie et al., 2009), making it valuable for modeling nonlinear rainfall thresholds (Fig. 2.6c). In our study, we used Scikit-learn's `QuadraticDiscriminantAnalysis` module with customized settings, including `priors` [0.3, 0.7] for class probabilities and `reg_param` 0.7 for regularization control.

Random Forest (RF)

Random Forest (RF) is a machine learning algorithm that improves accuracy by combining predictions from multiple decision trees, each trained on different data subsets (bootstrapping) (Breiman, 2001). Decision trees, used for classification and regression, are flowchart-like structures (Breiman et al., 1984).

In our study, we used Scikit-learn's `RandomForestClassifier` module with hyperparameter settings for optimization. For instance, `n_estimators` were set to 10 to limit the number of trees, criterion used entropy-based information gain, and hyperparameters like `max_depth`, `min_samples_split, and min_samples_leaf` was adjusted for tree control and preventing overfitting. To address class imbalance, `class_weight` was set to {0:1, 1:18.7} to assign higher weights to the minority class (Pedregosa et al., 2011).

RF classifies data into 'landslide' or 'no-landslide' without providing a specific rainfall threshold for separating critical and non-critical events in the ID plane (Fig. 2.6d).

XGBoost

Extreme Gradient Boosting (XGBoost) is a powerful machine learning algorithm that improves accuracy by combining predictions from multiple decision trees through gradient boosting (Chen and Guestrin, 2016). Like Random Forest, XGBoost builds multiple decision trees, but it uses gradient boosting to enhance performance. This technique corrects errors from previous trees, leading to higher accuracy and better model performance (Friedman, 2001). XGBoost excels with large datasets and numerous input variables, making it a widely used algorithm for predictive modeling.

In our analysis, we employed the XGBoost algorithm using the `XGBClassifier` module from the `xgboost` Python package. We optimized the model by adjusting hyperparameters such as booster (`gbtree`), `max_depth` (6), and `scale_pos_weight` (18.7) to enhance its ability to capture complex data relationships.

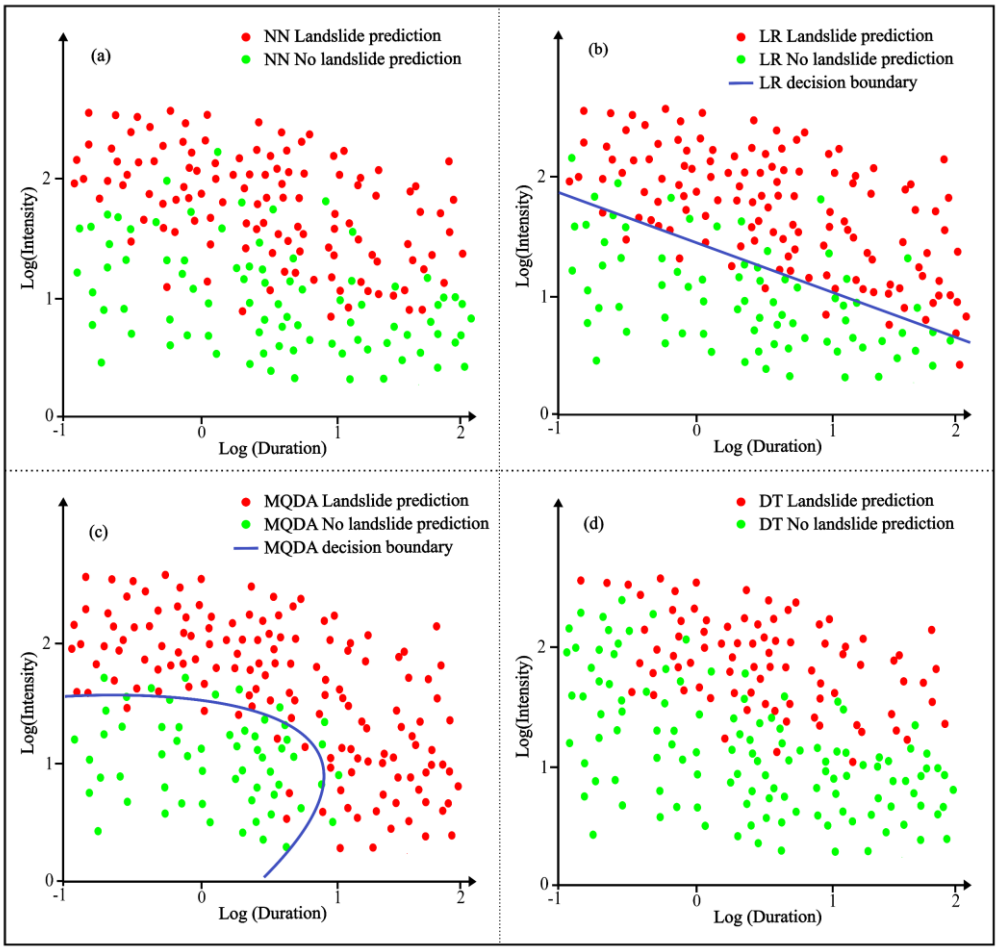


Figure 2.6. Differences in Outputs from Machine Learning algorithms for rainfall threshold evaluation. (**a**) Neural Network: this approach uses multiclass analysis for the individual classification of each rainfall event, lacking a distinct rainfall threshold. (**b**) Logistic Regression: the decision boundary can be depicted as a line in the 2D plot. Note that a critical rainfall event may appear below this line, influenced by other parameters not represented in this 2D visualization. (**c**) Multiparametric Quadratic Discriminant Analysis (MQDA): the decision boundary can be shown as a curve in the 2D plot. Similar to Logistic Regression, the impact of additional parameters is not visible in this plot. (**d**) RF, XGBoost, and TPOT: these methods employ multiclass Decision Tree (DT) approaches, which do not have clearly defined boundaries in their outputs.

Tree-Based Pipeline Optimization Tool (TPOT)

TPOT (Tree-Based Pipeline Optimization Tool) is not a model but an automated machine learning library designed to optimize decision tree-based models (TPOTClassifier) (Olson et al., 2016). TPOT employs genetic programming for autonomous model optimization (Olson & Moore, 2019). It evolves through generations of randomly generated models, converging on an optimized decision tree-based pipeline.

In our study, TPOT was configured with specific parameters: 150 `generations`, a `population_size` of 150, custom F2 score, and `early_stop` after 15 generations. This involved evaluating 22,500 pipeline configurations. It's crucial to note that TPOT operates as an automated optimization tool, providing limited user control beyond initial parameter settings, distinguishing it from traditional machine learning algorithms.

The optimal pipeline identified by TPOT is shown below:

```
model = make_pipeline(
    StackingEstimator(estimator=RandomForestClassifier(
    bootstrap=True, criterion="entropy",  max_features=0.1,
    min_samples_leaf=15,  min_samples_split=3, n_estimators=100)),
    StackingEstimator(estimator=MLPClassifier(alpha=0.1
    ,learning_rate_init=0.1)), SelectFwe(score_func=f_classif,
    alpha=0.032),
    StackingEstimator(estimator=DecisionTreeClassifier(
    criterion="entropy", max_depth=1, min_samples_leaf=13,
    min_samples_split=17)), GaussianNB())
```

Considerations on Bayesian Network

Bayesian Network Machine Learning (BN-ML) techniques are powerful tools for modeling complex dependencies and providing insights into variable importance (Neapolitan, 2003). However, in this study, BN-ML was not included due to several reasons. Firstly, the complexity of determining the network structure and the computational cost of performing inference with large datasets make these techniques less practical for our analysis. Secondly, BN-ML techniques are less frequently represented in related literature, which limits their comparability with the more commonly used methods. Lastly, to maintain the focus and manageability of our study, we opted to use methods that are more straightforward to implement and have established performance metrics for our specific application.

2.5.3 Experiments and performance evaluation

In our analysis, the dataset was split into a training set and a test set in an 80/20 ratio (Alpaydin, 2010). Since the dataset is highly imbalanced with a ratio of 18:1 for the target variable (landslide events), we used stratified sampling to ensure that the training and validation sets have the same proportion of critical and non-critical events. Stratified sampling randomly selects samples from each group in a way that maintains the overall class balance (Kohavi, 1995).

Algorithms were then evaluated based on the four values of the confusion matrix (Bishop, 2006).

Among these, the F_β -score, a generalized form of the F-score, was specifically chosen for its ability to weigh recall and precision according to our research priorities. The general formula for the F_β-score is given by:

$$F_\beta = (1 + \beta^2) \cdot \frac{Precision \cdot Recalll}{(\beta^2 \cdot Precision) + Recall} \quad [2.4]$$

where:

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

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

For our specific application in landslide prediction, we prioritized recall over precision due to the critical nature of not missing actual landslide events. Thus, we chose a beta of 2, leading to the F2-score, which weighs recall higher than precision:

$$F2 = (1 + 2^2) \cdot \frac{Precision \cdot Recall}{(2^2 \cdot Precision) + Recall} \quad [2.5]$$

Among the many available performance metrics, we chose F2-score because it assigns more weight to false negatives (FN), which represent missed landslide events. Having a low number of missed alarms is of particular importance in landslide prediction. A high F2 score reflects a model's preference to issue false alarms (FP in this case) over missing alarms by significantly minimizing false negatives (FN). This scoring approach is particularly advantageous in situations where the costs of missing a true positive are much higher than those of handling false positives. The range of the F2 score is between 0 and 1, with a higher score indicating better performance.

The ROC AUC (Area Under the Receiver Operating Characteristic Curve) (Fawcett, 2006) was used to have a graphical representation of the algorithms performance based on the comparison between True Positive Rate on the y-axis, $TPR = TP/(TP + FN)$, and False Positive Rate on the x-axis, $FPR = FP/(FP + TN)$. By plotting the ROC curve, we can visually assess the model performance. A model with superior performance will exhibit a ROC curve that is closer to the top-left corner of the graph, indicating higher TPR and lower FPR.

Both metrics offer valuable insights into model performance, even in the presence of class imbalance.

The `roc_curve` and `roc_auc_score` functions from the scikit-learn Python library have been used to analyze the algorithms performance by means of the ROC curve. Traditional calibration and testing procedures were applied for conventional methods. For Machine Learning techniques, optimal performance was achieved through independent heuristic tuning of the algorithms' hyperparameters.

2.6 Results

2.6.1 Rainfall-landslide correlations

The landslide archive of the Emilia-Romagna region documents 6699 instances of landslides that occurred between 1939 and 2022 (Table 2.2, column 1). Out of these landslides, 4019 have a known triggering date with a daily precision or an uncertainty of fewer than seven days (column 4). The distribution of these landslides is not uniform across the eight alert areas, with around half of the occurrences happening in areas E1-E2. This unevenness is largely attributed to the unfavorable geological conditions in these regions, but it could also be due to more accurate detection of landslides made by local authorities.

The triggering rainfall for each of the 4019 documented landslides was manually identified (section 2.5.1). As mentioned earlier, this process was challenging, time-consuming, and involved a considerable degree of subjectivity due to the complexity of rainfall time series. Columns 5-6-7 in Table 2.2 summarize the results of this expert evaluation, indicating the number of triggering rainfalls identified with low uncertainty (type 1), moderate uncertainty (type 2), and those that could not be identified with the necessary objectivity for various reasons (type 3). The events suitable for the analysis are 2718 (type 1 + type 2) and account respectively for 39% (1532 landslides) and 30% (1186 landslides) of the 4019 historical records (column 9). On the other hand, those unsuitable for the analysis (type 3) account for 31% (1186 landslides). The fact that roughly one-third of the historical data is unsuitable for analysis represents a significant limitation that will be discussed later.

Critical rainfall only represents a small fraction of the overall rainfall events. During the 83 years of analysis, an average of 2000 rainfall events with varying intensity and duration were identified in each alert area using the D3P5 algorithm (section 2.5.1). Only a subset of these, ranging from 50 to 200, triggered landslides (Tab. 2.3). The percentage of critical rainfall events varied widely across the eight alert areas, ranging from 1.9% in area A1 to a maximum of 8.4% in areas E1 and E2, indicating that the dataset is heavily unbalanced. Further examination of the data reveals that while the number of rainfall events Is roughly similar in all eight areas, there are significant differences in both precipitation amounts and the number of landslides (Fig. 2.7). For instance, area E1, which experiences the highest annual rainfall (1596 mm/year), also has the highest number of landslides (761), whereas A2, which has much lower

precipitation levels (664 mm/year), has relatively fewer landslides (158). However, area E2, which receives a similar annual rainfall to A2 (651 mm/year), has the second-highest number of historical landslides (625). These disparities underscore the complex relationship between rainfall and landslides in the Emilia-Romagna region. This complexity is further underscored by the intensity-duration plot (Figure 2.8), which displays no distinct boundary between critical and non-critical rainfall events, making it challenging for any algorithm to find a distinct separation boundary.

To further understand the impact of various predictive features on landslide occurrence, we conducted a feature importance analysis using the machine learning models in this study. As illustrated in Figure 2.9, the analysis highlights that the total rainfall during the event is by far the most significant predictive feature. This was determined by calculating the feature importance based on the F2 score across 100 iterations for each method. The other parameters, while contributing positively, have relatively lower informative value, with their importance scores ranging between 0.0 and 0.1. This suggests that while total rainfall is a crucial factor in landslide prediction, other variables still play a role, albeit a smaller one.

Table 2.2. Overview of the Emilia-Romagna landslide dataset archive from 1939 to 2022. The table summarizes the precision of the date of landslide occurrence and the type of triggering rainfall across the eight alert areas. Type 1=low uncertainty; Type 2=Moderate uncertainty; Type 3=High uncertainty.

Area	Mean annual rainfall(mm)	Overall rainfall events (N)	Critical rainfall events (N)	Number of triggered landslides	Percentage of critical rainfall (%)
A1	954	2034	39	60	1.9
A2	664	1980	49	158	2.5
C1	1490	2537	115	276	4.5
C2	868	2474	106	330	4.3
E1	1596	2532	212	761	8.4
E2	651	1851	156	625	8.4
G1	1250	1954	95	245	4.9
G2	743	1817	96	263	5.3
Total	1027	17179	868	**2718**	**5.1**

Table 2.3. Overview of rainfall events detected by the automatic algorithm D3P5 during the 83-year analysis period. The table illustrates the average number of identified rainfall events, their intensity, and duration in each alert area. Notably, only a subset of these events, resulted in triggered landslides.

Alert area	Landslides triggered	Precision of the date of occurrence			Triggering rainfall				
		1 day	1 week	Less than 1 week	Type 1	Type 2	Type 3	No data	Type 1+2
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
A1	144	49	37	86	34	26	24	2	**60**
A2	494	125	136	261	95	63	72	31	**158**
C1	491	119	257	376	147	129	100	0	**276**
C2	734	141	372	513	216	114	183	0	**330**
E1	1761	426	611	1037	465	296	276	0	**761**
E2	1720	382	588	970	318	307	328	17	**625**
G1	523	111	211	322	132	113	77	0	**245**
G2	832	137	317	454	125	138	187	4	**263**
Total	**6699**	1490	2529	**4019**	1532	1186	1247	54	**2718**

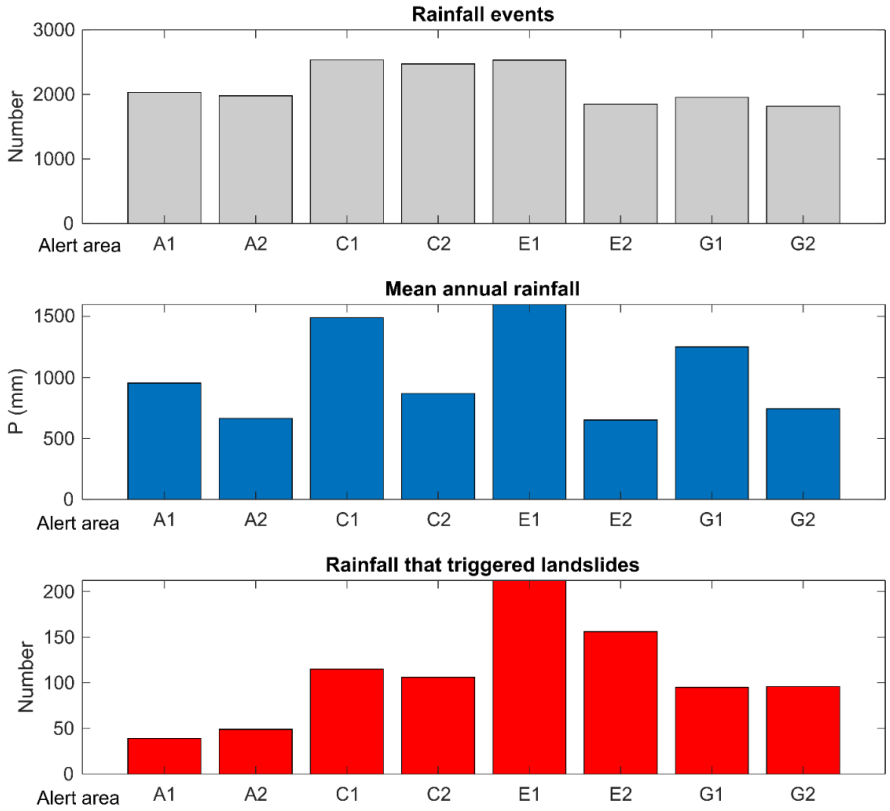


Figure 2.7. Variations in the total number of rainfall events (upper), mean annual precipitation (middle) and number of triggered landslides (lower) across the eight alert areas of the Emilia-Romagna region in the period 1939-2022.

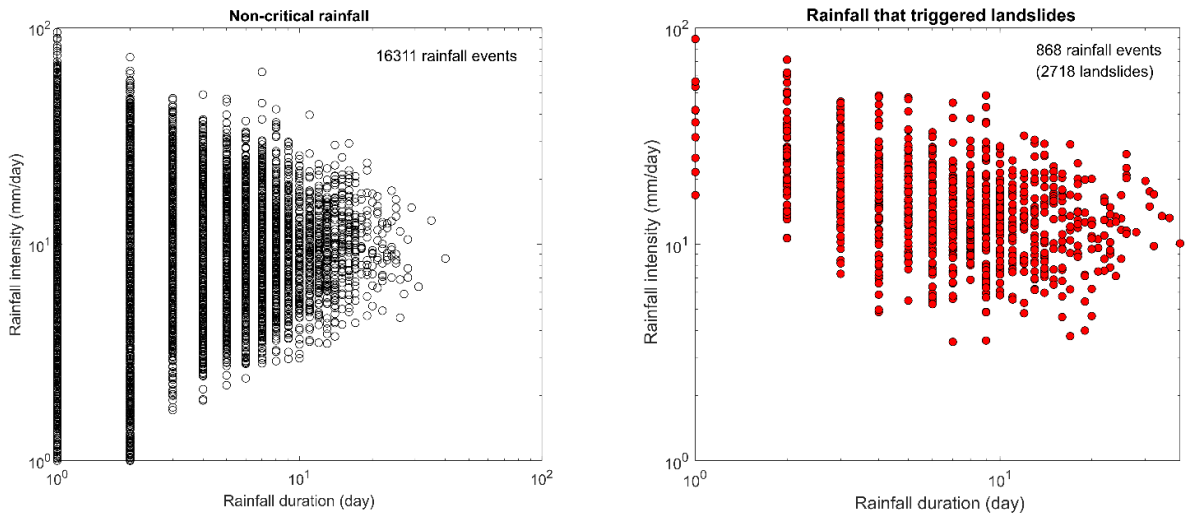


Figure 2.8. Log duration-log intensity charts showing the distribution of the total rainfall events occurred in the Emilia-Romagna region in the period 1939-2022 (left) and the critical rainfall that resulted in landslides (right).

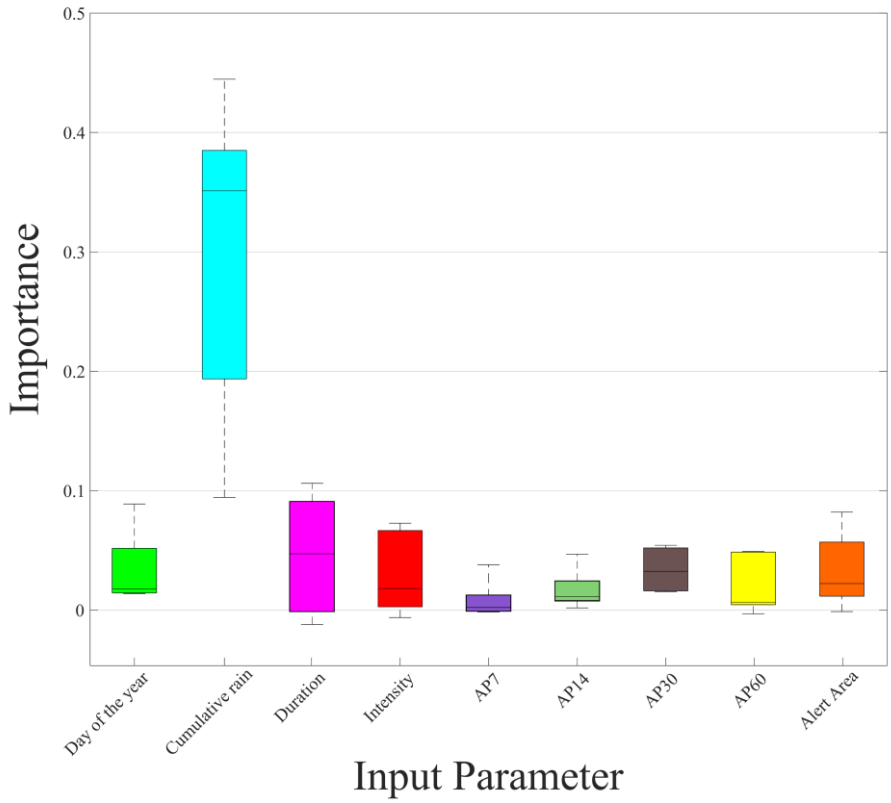


Figure 2.9. Feature importance analysis for machine learning models: This boxplot illustrates the feature importances for all the machine learning methods used in this study, calculated based on the F2 score across 100 iterations for each method. The parameters AP7, AP14, AP30, and AP60 refer to the antecedent precipitation over 7, 14, 30, and 60 days, respectively.

2.6.2 Comparative evaluation of rainfall thresholds algorithms

In this section, we detail the results from the comparative analysis of algorithms discussed in section 2.5.2, focusing on their accuracy in forecasting landslides as presented in Table 2.4 and Figure 2.10, which include confusion matrices, F2 scores, precision scores, and ROC AUC scores, detailed in section 2.5.3.

Based on the results presented in Table 2.4, Figure 2.10, and Figure 2.13, we make the following condensed observations:

Conventional Methods (BART, Frequentist, Percentile, LDA)

- BART demonstrates the highest accuracy among conventional methods with the highest F2 score (0.477) and ROC AUC score (0.806). It predicts a triggering rainfall event accurately once in every 5 positive predictions reflecting a low precision (0.182).
- Frequentist and Percentile methods display lower efficiency mainly due to their higher rate of false positives.
- LDA, while performing similarly to BART in terms of triggering event predictions, shows slightly better precision (0.188) with fewer false positives.

Machine Learning Methods (MLDA, NN, LR, MQDA, RF, XGBoost, TPOT)

- MLDA and MQDA slightly outperform the conventional LDA with higher F2 scores (0.504 for MLDA and 0.483 for MQDA) and ROC AUC scores, indicating the limited effectiveness of the other parameters.
- Neural Networks (NN) and Random Forest (RF) exhibit similar F2 and ROC AUC scores, but with a significant difference in the number of false negatives handled.
- Logistic Regression (LR) exhibits a good F2 performance (0.522) with the lowest number of false positives, reflecting its precision in event prediction.
- XGBoost, the top-performing method, not only has the highest F2 score (0.531) and ROC AUC score (0.848) but also maintains good precision (0.223), accurately predicting a triggering rainfall event once in every four instances.
- TPOT achieved good results with the highest F2 (0.531) and precision score (0.266) among the methods tested. However, for our purposes, its reliability is compromised due to the high number of false negatives (51).

In summary, Machine Learning techniques generally outperform conventional methods as shown by their F2 scores, yet the high number of True Negatives in our unbalanced dataset skew ROC AUC scores, as illustrated in Fig. 2.11.

The observed low precision across models results directly from prioritizing F2 during training, which allows more false positives to reduce false negatives, reflecting our strategic focus on minimizing missed alarms.

Figure 2.12 provides a comparative analysis of seven Machine Learning techniques using 2D plots, which depict the probability that a rainfall event is classified as critical by each model within the two-dimensional space defined by rainfall intensity and duration during the training process. Given the multidimensional nature of these ML models, the plots serve only as approximations of how these variables influence classifications. This simplification allows us to visualize complex interactions in a more comprehensible form, illustrating how ML algorithms align with established threshold trends of conventional methods, albeit in a simplified manner that may not fully capture the models' capabilities. It's notable how the best-performing models, in terms of scores, exhibit distinct behaviors. TPOT, as shown in Figure 2.12g, assigns a binary output and, according to its confusion matrix, adopts a higher tolerance threshold compared to conventional methods. This results in better performance but also introduces higher risk. Conversely, XGBoost (Figure 2.12f) seems to maintain a more conservative profile. Both models classify a significant number of events as non-landslide which are well above conventional tolerance levels. This demonstrates how additional parameters influence decisions, leading to a decision-making process that resembles the BART threshold for short durations and high intensities, yet aligns with lower-slope thresholds for long-duration events.

Table 2.4. Performance evaluation of conventional methods (Bayes, Frequentist, Percentile, LDA) and Machine Learning techniques (NN, LR, MLDA, MQDA, RF, XGBoost, TPOT).

Method	Scores			Confusion Matrix	
	F2	Precision	ROC AUC	TN	FP
				FN	TP
BART	0.477	0.182	0.806	2631	631
				34	140
Frequentist	0.398	0.131	0.762	2328	934
				33	141
Percentile	0.384	0.133	0.738	2433	829
				47	127
LDA	0.465	0.188	0.783	2710	552
				46	128
NN	0.500	0.197	0.815	2677	582
				34	143
LR	0.522	0.221	0.822	2775	487
				36	138
MLDA	0.504	0.198	0.822	2681	581
				31	143
MQDA	0.483	0.183	0.811	2629	633
				32	142
RF	0.500	0.193	0.823	2654	608
				29	145
XGBoost	0.531	0.223	0.848	2772	490
				33	141
TPOT	0.531	0.266	0.801	2923	339
				51	123

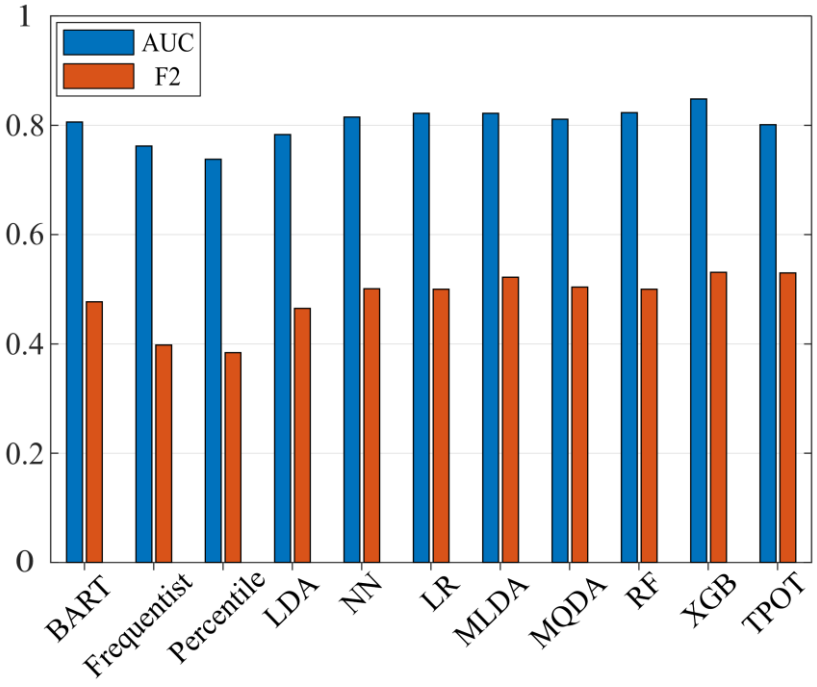


Figure 2.10. Distribution of ROC AUC and F2 scores for different predictive models. The x-axis represents the methods, and the y-axis displays the corresponding AUC, in blue bars, and F2, in red bars, scores.

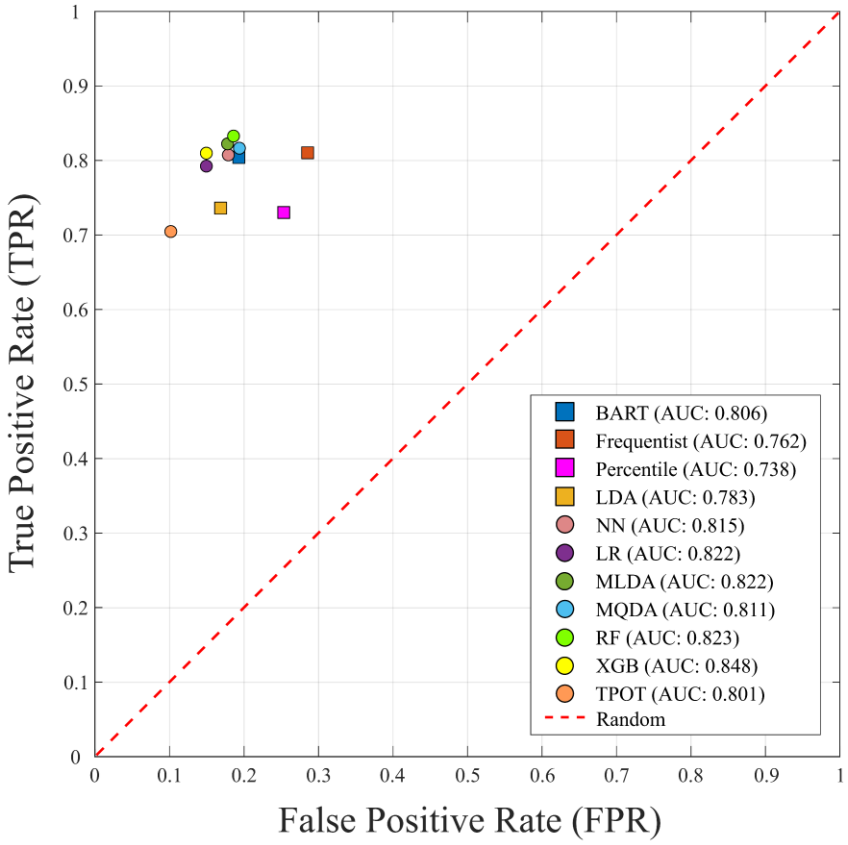


Figure 2.11. Comparison of ROC AUC optimal operating points for each method.

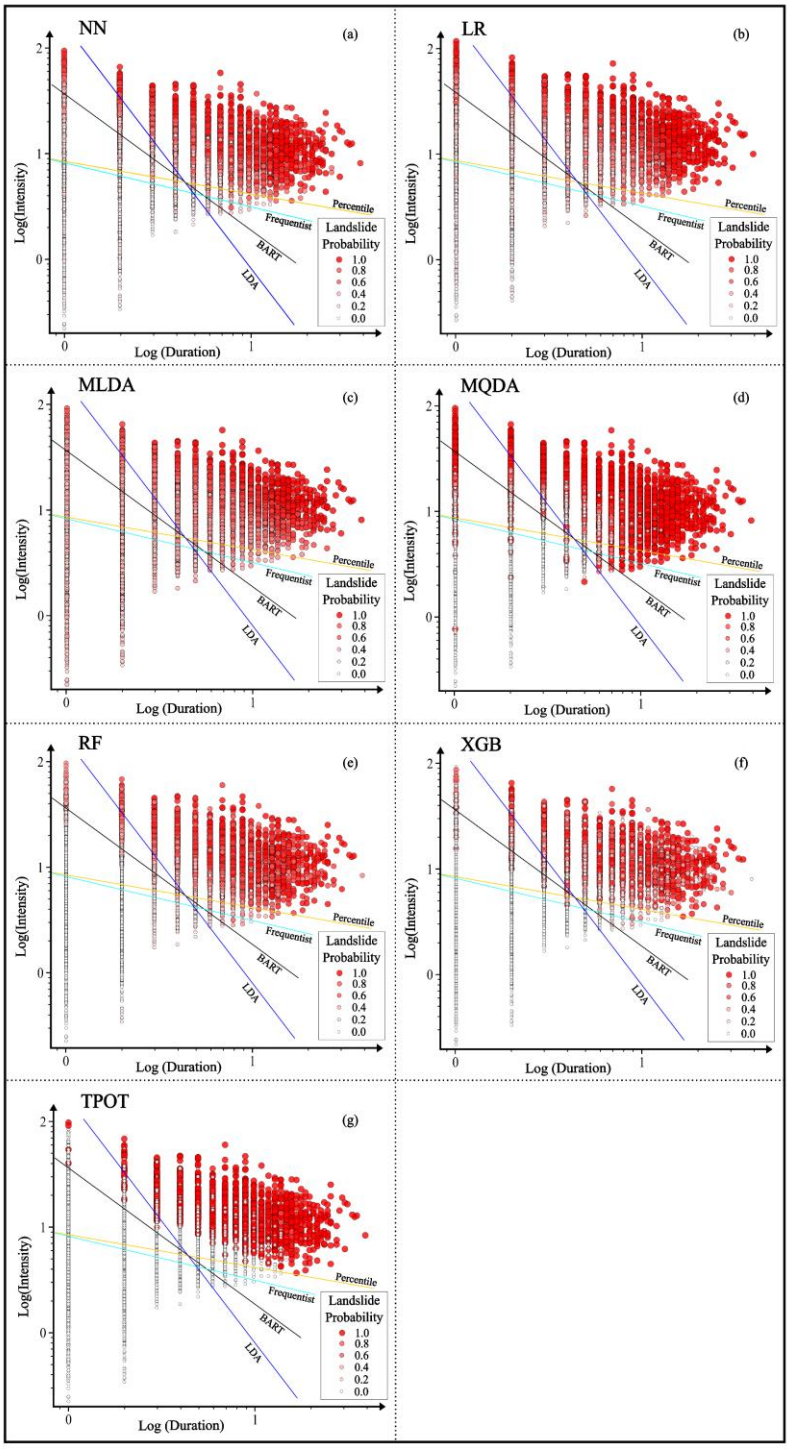


Figure 2.12. Comparison between conventional and Machine Learning methods in the rainfall Intensity-Duration plot. Rainfall events are denoted as points, with their color varying from white to red, indicating the probability of being classified as critical (on a scale from 0 to 1) by the different ML algorithms during their train process. Lines indicate the optimal rainfall thresholds obtained by the conventional methods. The equations for the optimal rainfall thresholds in the conventional methods are as follows: BART in black ($I = -1 \cdot \log(D) + 1.7347$), LDA in blue ($I = -1.8602 \cdot \log(D) + 2.3542$), Frequentist approach in light blue ($I = -0.3159 \cdot \log(D) + 1.2504$), and Percentile method in orange ($I = -0.2341 \cdot \log(D) + 1.243$).

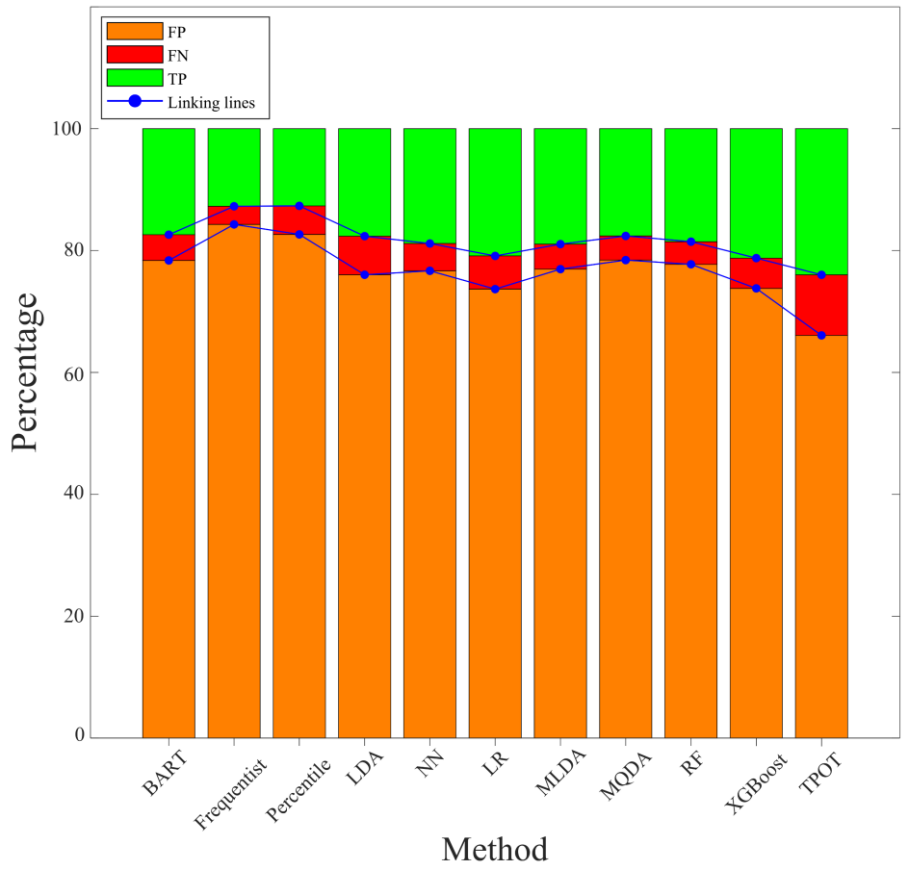


Figure 2.13. Normalized Confusion Matrix Components (Excluding True Negatives) Comparison. This figure presents the normalized percentages of False Positives (FP), False Negatives (FN), and True Positives (TP) for the methods investigated in this study. The connecting lines illustrate the trends and comparative performance of the methods across the different components.

2.7 Discussion

2.7.1 Performance and usability of Machine Learning methods

The results from our study demonstrate that machine learning methods, particularly XGBoost, have outperformed conventional methods such as BART, which is currently employed as a decision-support tool within the Civil Protection Agency of the Emilia-Romagna Region (Italy) (Berti et al., 2015). For instance, when comparing the performance of XGBoost with BART, a reduction in the number of false positives—from 631 in BART to 490 in XGBoost—is observed as detailed in Table 2.4. Moreover, the variables used in this study for the ML methods are readily available in real-time, making the integration feasible and potentially beneficial.

To further operationalize these findings, a decision-support platform integrating the improved rainfall thresholds and machine learning predictions was developed as part of this research. This platform, described in detail in Chapter 6, is designed to assist emergency management agencies

in real-time decision-making by leveraging the predictive capabilities demonstrated in this study.

In scenarios where warnings do not lead to costly actions, such as evacuations or road closures, as is the case with the Civil Protection Agency of the Emilia-Romagna Region, as outlined in the Regional Council Decree No. 1761 (dated November 30, 2020), a certain level of false positives may be considered acceptable. In some cases, indeed, tolerating a higher rate of False Positives might even be preferable, as it could alleviate the burden of responsibility on risk managers in the event of unforeseen events. Therefore, while our improvement in terms of reducing false positives may not be decisive, it remains an enhancement.

Beyond mathematical scores, other aspects need consideration, particularly in terms of method communication and practicality. Conventional methods, with their threshold approach in a 2D plane, offer ease of use in practical settings through a simple, visual boundary line. In contrast, while Machine Learning techniques are potentially more powerful, they vary significantly in their levels of interpretability. While some ML algorithms, like neural networks and deep learning, often function as 'black boxes' with limited visibility into their decision-making processes, many others, such as decision trees, linear models, and Bayesian networks, are inherently more interpretable and provide greater insight into their workings. This variability in transparency can be a challenge as users may need to trust the model without direct control over the prediction mechanism. However, the level of this challenge depends on the specific ML technique employed. Although complexity and difficulty in usage might be acceptable with superior results, in our case, the performance improvement of Machine Learning methods over traditional techniques does not seem sufficiently clear-cut to universally justify these challenges.

2.7.2 Efficacy of multiparametric methods

Multiparametric models, including those based on Machine Learning, can capture the complexity of real-world problems, often surpassing simpler models in effectiveness. These methods, by integrating a broader range of variables, have the potential to deliver more precise and detailed predictions. They excel at identifying intricate patterns and relationships, as well as managing nonlinear interactions among variables – abilities that are particularly valuable in the complex domain of predicting rainfall-induced landslides. However, the observation that multiparametric methods do not significantly outperform simple two-parameter models in our study is somewhat unexpected.

In our analysis, the seven Machine Learning methods were applied using seven additional parameters beyond the conventional metrics of rainfall duration and intensity, including factors like seasonality and antecedent rainfall. Despite incorporating these extra parameters, the improvement in predictive performance was modest. To further explore this, we also ran the

Machine Learning algorithms using only rainfall intensity and duration, paralleling the approach of traditional models. The results, as detailed in Table 2.5, reveal that Machine Learning algorithms, when constrained to this simpler framework, scored lower in both F2 and ROC AUC metrics compared to their multiparametric versions.

This outcome underscores that the true strength of Machine Learning lies in its capacity to process and leverage complex, multi-variable datasets. Consequently, the limited effectiveness of ML in our study implies that the extra parameters included were not particularly impactful. This underscores a crucial, albeit intuitive, principle: the quality and relevance of input parameters are more crucial than the model's complexity in enhancing the predictive accuracy of rainfall thresholds. In our specific instance, although the additional parameters used in the multiparametric methods were chosen with care, they did not significantly enhance the distinction between critical and non-critical rainfall events. On the other hand, slope stability is influenced by numerous factors beyond rainfall, such as stress history, soil moisture, subsurface flows, human activities, changes in vegetation, and others, making it a challenge to pinpoint the most relevant factors when working at regional scale.

Machine Learning methods, however, now offer unprecedented capabilities to analyze large arrays of interrelated factors efficiently. This capability encourages the collection and analysis of a broader range of factors more comprehensively than ever before, potentially leading to more insightful and accurate models in complex scenarios like rainfall-induced landslide prediction.

Table 2.5. Influence of the number of parameters on the prediction capability. The table compares the F2 and AUC scores obtained by the conventional methods (gray cells) with those obtained by the Machine Learning methods executed with different number of input parameters.

Method	Score			
	2 parameters (I, D)		9 parameters (I, D, dY, C, A7, A14, A30, A60, Ac)	
	F2	AUC	F2	AUC
BART	0.477	0.806	/	/
Frequentist	0.398	0.762	/	/
Percentile	0.384	0.738	/	/
LDA	0.465	0.783	/	/
NN	0.450	0.789	0.518	0.814
LR	0.468	0.79	0.532	0.829
MLDA	/	/	0.522	0.834
MQDA	0.437	0.79	0.501	0.83
RF	0.451	0.793	0.528	0.851
XGBoost	0.455	0.794	0.604	0.856
TPOT	0.451	0.749	0.542	0.791

2.7.3 Enhancing the reliability of the rainfall threshold

To improve the predictive capability of the rainfall threshold, it is essential to go beyond the mere computational power of the algorithm. While more advanced methodologies, such as Machine Learning, offer powerful tools for predictions, their effectiveness can be compromised by several significant sources of uncertainty. These uncertainties contribute to the overlap of

critical and non-critical rainfalls shown in Figure 2.8, emphasizing the complexity of the prediction task. Key sources of data uncertainty include:

Initiation Date Ambiguity: The lack of precision in pinpointing the exact date of a landslide's onset leads to variability in the dataset, primarily because it complicates the identification of the triggering rainfall event. This ambiguity often stems from the challenge of recognizing the true beginning of the landslide, as it is typically recorded in chronicles only upon being noticed.

Variable Response to Rainfall: The way landslides react to similar rainfall events varies, influenced by their distinct geological, geomorphological, and hydrogeological attributes, along with their stress history and any human interventions. When defining rainfall thresholds at a regional scale, it's impractical to isolate and analyze each of these factors separately, leading to an inherent increase in uncertainty.

Definition of the Rainfall Event: The methodology employed in defining a rainfall event significantly influences its determined duration and mean intensity. Currently, the absence of established guidelines for selecting this criterion results in its definition being highly subjective. The choice of parameters, such as the specific start and end times of the rainfall, the threshold values for intensity, and the inclusion of antecedent moisture conditions, greatly affects the characterization of the event. This variability in definition can lead to diverse interpretations and assessments of the same rainfall data.

In summary, enhancing the predictive capacity of the rainfall threshold necessitates an investment in data quality. This involves a more detailed characterization of the outcomes following a rainfall event. While historical data may be limited in this regard, future efforts could focus on improving data quality, allowing Machine Learning methods to fully deploy their potential.

2.8. Conclusions

This study leads to several key insights:

- Identifying rainfall thresholds in regions with intricate geological features, such as the Emilia-Romagna Region in Italy, is notably challenging. This complexity primarily arises from the difficulty in differentiating between rainfall events that are critical for landslide initiation and those that are not. These events often overlap significantly on the rainfall duration-intensity chart, making clear differentiation a complex task.
- In our analysis, we observed significant variations in the threshold equations obtained by different conventional methodologies, particularly when comparing early machine learning approaches (like BART and LDA) to traditional mathematical-statistical methods (such as Frequentist and Percentile approaches).

- Machine learning techniques, especially the eXtreme Gradient Boosting (XGBoost) algorithm, demonstrated superior performance over conventional two-parameter methods. Their enhanced effectiveness is reflected in both improved prediction scores and a significant reduction in False Alarms.
- The boost in predictive accuracy provided by machine learning methods was however relatively modest, considering that these methods utilized eight additional rainfall parameters.
- The marginal impact of these additional parameters implies that their informational value, particularly concerning antecedent rainfall over periods of 7 to 60 days, is somewhat limited when compared to more direct factors like rainfall intensity or cumulative duration.
- Progress in predictive modeling for landslide risks in the future should focus on improving the quality and specificity of data. Developing a dataset that is more detailed and informative has the potential to create models that are both more accurate and reliable, offering greater benefits than the mere application of advanced methods.

In summary, this study underscores the ongoing challenges faced in accurately predicting landslides caused by rainfall in complex geological conditions. It brings to light the continuous necessity for improvements in both the quality and volume of data, especially in the current context where advanced artificial intelligence techniques have become widely accessible. This emphasis on data refinement is crucial because the effectiveness of these powerful AI methods is highly dependent on the quality and depth of the dataset they utilize. Therefore, as we advance in employing sophisticated AI tools for environmental analysis and prediction, equal emphasis must be placed on enhancing the data that forms the foundation of these predictive models.

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) and from Fondazi one Cassa di Risparmio in Bologna (Fondazione CARISBO).

Funding

Open access funding provided by Alma Mater Studiorum – Università di Bologna within the CRUI-CARE Agreement.

Data Availability

The authors are pleased to make the dataset used in this study available to interested researchers upon request. Please feel free to reach out to the corresponding author for access to the data.

Code availability

The authors are pleased to make the software code utilized in this experiment available to interested researchers upon request. Please feel free to reach out to the lead author for access to the code.

2.9 References

Aleotti, P. (2004). A warning system for rainfall-induced shallow failures, *Eng. Geol.*, 73, 247–265. doi:10.1016/j.enggeo.2004.01.007

Aleotti, P., and R. Chowdhury (1999). Landslide hazard assessment: Summary review and new perspectives, *Bull. Eng. Geol. Environ.*, 58, 21–44. doi:10.1007/s100640050066

Alpaydin, E. (2010). Introduction to Machine Learning, 2nd edn. *MIT Press*, Cambridge, MA.

Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface*, 117(F4). doi:10.1029/2012JF002367

Berti, M., Martina, M.L.V., Franceschini, S., Pignone, S., Simoni, A., Pizziolo, M. (2015). Implementation of a Probabilistic Model of Landslide Occurrence on a Civil Protection Alert System at Regional Scale. In: Lollino, G., et al. *Engineering Geology for Society and Territory – Volume 2. Springer*, Cham, doi:10.1007/978-3- 319-09057-3_110.

Bertolini, G., Guida, M. & Pizziolo, M. (2005). Landslides in Emilia-Romagna region (Italy): strategies for hazard assessment and risk management. *Landslides* 2, 302–312. doi:10.1007/s10346-005-0020-1

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (1st Edition). Springer.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324

Breiman L., Jerome Friedman, Charles J. Stone, & R.A. Olshen. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC. doi:10.1201/9781315139470

Brown, C. E. (1998). Multiple Logistic Regression. *Applied Multivariate Statistics in Geohydrology and Related Sciences*, 147–151. doi:10.1007/978-3-642-80328-4_12

Brunetti MT, Peruccacci S, Rossi M, Luciani S, Valigi D, Guzzetti F (2010). Rainfall thresholds for the possible occurrence of landslides in Italy. *Nat Hazards Earth Syst Sci* 10:447–458. https://doi.org/10.5194/nhess-10-447-2010

Caine, N. (1980). The rainfall intensity-duration control of shallow landslides and debris flows, *Geogr. Ann., Ser. A*, 62(1–2), 23–27. doi:10.2307/520449

Calvello, M., & Piciullo, L. (2016). Assessing the performance of regional landslide early warning models: The EduMaP method. *Natural Hazards and Earth System Sciences*, 16(1), 103–122. doi:10.5194/nhessd-3-6021-2015

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794. doi:10.1145/2939672.2939785

Corominas, J. (2000). Landslides and climate, in *Proceedings of the 8th International Symposium on Landslides*, edited by E. Bromhead, N.Dixon, and M. L. Ibsen, pp. 1–33, A. A. Balkema, Cardiff, Wales.

Cox D. R., The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 20, Issue 2, July 1958, Pages 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Crosta GB, Frattini P (2001). Rainfall thresholds for triggering soil slips and debris flow. In: *Proc. Of the 2nd EGS Plinius Conference on Mediterranean Storms: Publication CNR GNDCI*, 2547, pp. 463–487.

De Vita, P., Reichenbach, P., Bathurst, J. C., Borga, M., Crozier, G. M., Glade, T., Guzzetti, F., Hansen, A., & Wasowski, J. (1998). Rainfall-triggered landslides: a reference list. *Environmental Geology* 1998 35:2, 35(2), 219–233. doi:10.1007/S002540050308

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. doi:10.1016/J.PATREC.2005.10.010

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188. doi:10.1111/J.1469-1809.1936.TB02137.X

Friedman J.H. (2001). Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451

Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, 162, 227–252. doi:10.1016/J.EARSCIREV.2016.08.011

Glade, T., Crozier, M., & Smith, P. (2000). Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "Antecedent Daily Rainfall Model." *Pure and Applied Geophysics*, 157 (6–8),1059–1079. doi:10.1007/S000240050017

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press*, Cambridge, MA. Available online: https://www.deeplearningbook.org/ (Accessed on 26 June 2007).

Güemes-Peña, D., López-Nozal, C., Marticorena-Sánchez, R. et al. (2018). Emerging topics in mining software repositories. *Prog Artif Intell* 7, 237–247. https://doi.org/10.1007/s13748-018-0147-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics 2007*, 98(3), 239–267. doi:10.1007/S00703-007-0262-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007b). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3–4), 239–267. doi:10.1007/s00703-007-0262-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity-duration control of shallow landslides and debris flows: An update. *Landslides*, 5(1), 3–17. doi:10.1007/S10346-007-0112-1

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Stanford, CA: Stanford University. doi:10.1007/978-0-387-84858-7

Huang, L., Xiang, Ly. (2018). Method for Meteorological Early Warning of Precipitation-Induced Landslides Based on Deep Neural Network. *Neural Process Lett 48*, 1243–1260. doi:10.1007/s11063-017-9778-0

Hu, X., Wu, S., Zhang, G., Zheng, W., Liu, C., He, C., ... & Zhang, H. (2021). Landslide displacement prediction using kinematics-based random forests method: A case study in Jinping Reservoir Area, China. *Engineering Geology*, 283, 105975. doi:10.1016/j.enggeo.2020.105975

Kohavi R (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence – Volume 2*. Morgan Kaufmann Publishers Inc., San Francisco, pp 1137–1143. https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf (accessed 12 Dec 2023)

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. doi:10.1038/nature14539

Liu, Z., Gilbert, G., Cepeda, J. M., Lysdahl, A. O. K., Piciullo, L., Hefre, H., & Lacasse, S. (2021). Modelling of shallow landslides with machine learning algorithms. *Geoscience Frontiers*, 12(1), 385–393. doi:10.1016/J.GSF.2020.04.014

Mondini, A.C., Guzzetti, F. & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nat Commun* 14, 2466. doi:10.1038/s41467-023-38135-y

Moore, D.S. and McCabe, G.P. (1993). Introduction to the Practice of Statistics. WH Freeman/Times Books/Henry Holt & Co., New York.

Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.

Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science. *GECCO 2016 - Proceedings of the 2016 Genetic and Evolutionary Computation Conference*, 485–492. doi:10.1145/2908812.2908918

Olson, R.S., Moore, J.H. (2019). TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds) Automated Machine Learning. *The Springer Series on Challenges in Machine Learning*. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_8

Orland, E., Roering, J. J., Thomas, M. A., & Mirus, B. B. (2020). Deep learning as a tool to forecast hydrologic response for landslide-prone hillslopes. *Geophysical Research Letters*, 47. doi:10.1029/2020GL088731

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. doi:10.48550/arXiv.1201.0490

Ramos-Cañón, A.M., Prada-Sarmiento, L.F., Trujillo-Vela, M.G. et al. (2016). Linear discriminant analysis to describe the relationship between rainfall and landslides in Bogotá, Colombia. *Landslides 13*, 671–681. doi:10.1007/s10346-015-0593-2

Rossi F.; Witt A.; Guzzetti F.; Malamud B.D.; Peruccacci S., (2010). Analysis of historical landslide time series in the Emilia-Romagna Region, Northern Italy., *Earth surface processes and landforms (Print)* 35 (2010): 1123–1137.

Saito, H., D. Nakayama, and H. Matsuyama (2010). Relationship between the initiation of a shallow landslide and rainfall intensity—Duration thresholds in Japan, *Geomorphology*, 118, 167–175. doi:10.1016/j.geomorph.2009.12.016

Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003

Segoni, S., Piciullo, L., & Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15(8), 1483–1501. doi:10.1007/S10346-018-0966-4

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. doi:10.1201/9781315140919

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. doi:10.1007/978-0-387-21706-2

Wang, G., & Sassa, K. (2003). Pore-pressure generation and movement of rainfall-induced landslides: effects of grain size and fine-particle content. *Engineering geology*, 69(1-2), 109-125. doi:10.1016/S0013-7952(02)00268-5

Wieczorek, G. F., & Glade, T. (2007). Climatic factors influencing occurrence of debris flows. *Debris-Flow Hazards and Related Phenomena*, 325–362. doi:10.1007/3-540-27129-5_14

Zhao, Y., Meng, X., Qi, T., Li, Y., Chen, G., Yue, D., & Qing, F. (2022). AI-based rainfall prediction model for debris flows. *Engineering Geology*, 296, 106456. doi:10.1016/j.enggeo.2021.106456

Chapter 3

RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event

3.1 Preface

This chapter explores a significant event in hydrogeological risk management: the comprehensive mapping and analysis of the May 2023 landslide event in the Emilia-Romagna region. Following the extreme weather events of May 2023, this work became necessary to map and assess the landslides triggered by intense rainfall. The project was carried out in collaboration with the Emilia-Romagna regional authorities, the University of Modena and Reggio Emilia, and the Po River District Authority.

The dataset presented here, compiled through a coordinated effort involving manual mapping with high-resolution aerial imagery and advanced geospatial techniques, represents a key tool for both immediate recovery efforts and long-term hazard assessment. The resulting inventory, which includes over 80,000 landslides, provides an unprecedented level of detail and serves as a crucial resource for refining future landslide susceptibility models in the region.

Although my role was focused on specific aspects of the research, this work is the result of a larger, multi-institutional collaboration. My contributions formed part of this coordinated effort, which has been instrumental in supporting ongoing research aimed at mitigating the impact of future hydrogeological events in the Emilia-Romagna region.

The chapter is based on a paper currently under review in the journal *Earth System Science Data (ESSD):*

Berti M.[1], Pizziolo M.[2], Scaroni M.[2], Generali M.[2], Critelli V.[3], Mulas M.[3], Tondo M.[3], Lelli F.[3], Fabbiani C.[3], Ronchetti F.[3], Ciccarese G.[1], Dal Seno N.[1], Ioriatti E.[1], Rani R.[1], Zuccarini A.[1], Simonelli T.[4], Corsini A.[2]. RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event [preprint], https://doi.org/10.5194/essd-2024-407, 2024.

[1] Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy.

[2] Regione Emilia-Romagna, Area Geologia, Suoli e Sismica, Bologna, Italy.

[3] Department of Chemical and Geological Sciences, University of Modena and Reggio-Emilia, Modena, Italy.

[4] Autorità di Bacino Distrettuale del Fiume Po, Parma, Italy.

3.2 Abstract

Landslide inventories are crucial for evaluating susceptibility, hazards, and risks, and for devising resilience strategies in mountainous regions. This importance is amplified in the context of climate change, as existing inventories might not adequately reflect changing stability conditions. In May 2023, the Emilia-Romagna region of Italy was hit by two major rainfall events, leading to widespread flooding and the triggering of thousands of landslides. Predominantly, these were shallow debris slides and debris flows, occurring on slopes previously deemed stable based on historical data with no prior landslides recorded. Our team supported the Civil Protection Agency through field surveys and mapping efforts to pinpoint and record these landslides, prioritizing areas critical to immediate public safety and focusing on thorough mapping for future recovery planning. The outcome is a detailed map of all landslides induced by these events, manually identified using high-resolution aerial photography (0.2 m pixel resolution, RGB+NIR four bands) and categorized with the help of a 3D viewer. This comprehensive landslide inventory, comprising 80997 polygons, has been made openly accessible to the scientific community.

3.3 Introduction

Landslide inventories are crucial for susceptibility, hazard, and risk assessments and management (Soaters & Van Westen, 1996; Fell et al., 2008; Galli et al., 2008; Corominas et al., 2014). In Europe, landslides inventories are compiled on a national to a regional basis (Van Den Eeckhaut and Hervás, 2012) and can be supported by advanced landslides recognition and monitoring techniques (Guzzetti et al., 2012; Jaboyedoff et al., 2012; Amatya, 2021; Catani, 2021; Bhuyan et al., 2023). Landslides inventories should be as complete and spatially accurate as possible and, also, they should consistently distinguish and classify different landslides types. These factors are important for improving frequency-area analyses (Malamud et al., 2004) and for obtaining reliable statistically-based landslides susceptibility maps, thanks to complete input data (Steger et al., 2017; Gaidzik et al., 2021) and disjunct analysis of landslides types (Zêzere, 2002).

Generally, inventories of large-scale landslides are quite complete, since their geomorphic features that remain evident long after their occurrence and they can have slow movements detectable by remote sensing (Bertolini et al., 2017; Rosi et al., 2018; Luetzenburg et al., 2022; Ardizzone et al., 2023). On the contrary, regional or national inventories might not include a complete record of past shallow rainfall induced landslides, unless they have been mapped soon after occurrence, i.e. before becoming hardly recognizable due to vegetation growth, rill erosion, or land cultivation (Guzzetti et al., 2004; Crozier, 2005; Cardinali et al., 2006; Zieher et al., 2016; Hao et al., 2020, Santangelo et al., 2023). Therefore, it is important that existing national and regional landslides inventories are updated after each specific intense rainfall

event, so to collect data that are essential, also, for a more systematic usage of landslide susceptibility maps in land-use planning (Fell et al., 2008). This is a challenging and important task, as the incidence of shallow rainfall induced landslides is likely to increase in Europe due to climate change (Gariano and Guzzetti, 2016; Handwerger et al., 2022; Auflič et al., 2023).

The Emilia-Romagna Region (Northern Italy), land-use planning and land-use restrictions are based on an inventory map of landslides at 1:10.000 scale (Bertolini et al., 2005) and on a catalogue of thousands of records referring to the activation or reactivation of landslides in the past (Piacentini et al., 2018). These documents are updated after every occurrence or reactivation of large-scale landslides and after multiple occurrences of rainfall-induced shallow landslides. In recent years, updates have been necessary to include hundreds of debris flows triggered during the rainstorm events that hit Parma province in September 2014 (Corsini et al., 2017) and Piacenza province in October 2015 (Scorpio et al., 2018; Ciccarese et al., 2020). In May 2023, the entire southern sector of Emilia-Romagna (from Rimini to Reggio Emilia provinces) has been hit by two consecutive exceptional rainfall events that triggered thousands of first-failure landslides. A screening of these landslides has been rapidly provided by Ferrario et al. (2024) and Notti et al. (2024) by using supervised and unsupervised satellite-based identification methods. However, these datasets did not aim to, and did not reach, the level of completeness, consistency and accuracy required for updating the official landslide inventory that, in turn, determines significant practical consequences for the management of the mapped areas and their surroundings.

In this paper, we present the landslide inventory dataset of the May 2023 Emilia-Romagna events which has been designated as reference map by the Emilia-Romagna Region and the Po River Authority for the “Special Plan for interventions against situations of hydrogeological instability” (approved in preliminary version in April 2024), to aid the Commission for Reconstruction in implementing the recovery phase. This spatial dataset is based on expert-based identification, mapping and classification of landslides on high-resolution aerial images taken shortly after the second event (0.2 m resolution, RGB and near-infrared). Particular attention has been given to the consistency of landslides type classification, which required development and application of an algorithm for data harmonization across areas surveyed by different operators.

3.4 The May 2023 Emilia-Romagna event

The Emilia-Romagna region is located in northern Italy and stretches from the Apennine Mountains to the Po River Valley and eastward to the Adriatic Sea (Fig. 3.1a). It is one of Italy's most economically prosperous regions, with a strong industrial base in automotive, machinery, food processing, and ceramics. The Po River Valley plays a vital role in agriculture, while the eastern coastline is a hub for both domestic and international tourism. By contrast, the Apennine

Mountains present a more subdued economic landscape. Economic activity and population in these mountains have declined since the 1960s, and now focus on agrotourism, ecotourism, and niche markets. Regional initiatives are underway to foster economic growth in these mountainous areas.

In May 2023, the Emilia-Romagna region was struck by two exceptional rainfall events. The first, from May 1-3, delivered approximately 200 mm of rain over a span of 48 hours. Only two weeks later, on May 16-17, a second event matched this intensity, with rainfall totals reaching 200-250 mm within another 48-hour window. The recurrence interval for a single two-day event was estimated to exceed 100-300 years, but the combined effect of these two closely timed events far surpassed 500 years (Brath et al., 2023). Both events impacted roughly the same area in the eastern part of the region (Fig. 3.1).

These rainfalls led to extensive flooding across the Po plain and triggered thousands of landslides in the Apennines. The total damages have been estimated to surpass 9 billion euros, affecting roads, railways, buildings, and cultural heritage sites, along with the destruction of bridges, power facilities, and communication lines. Additionally, agricultural fields, farming operations, and cultivated slopes saw significant disruption over an area of about 1000 km^2. Fifteen people lost their lives due to the flooding and two due to landslides.

The Emilia-Romagna region and the Italian Government promptly responded to the event mobilizing all necessary resources. The primary focus was on the Po plain area, which is densely populated and houses the majority of industrial and agricultural activities. Consequently, the severe issues caused by landslides in the mountainous regions were initially overlooked. Over time, the significance of these issues became apparent, but even a year after the disaster the situation remains critical. The impact of landslides in the Apennines has been especially severe due to the local economy's vulnerability, the extensive damages to infrastructure, and the significant land loss, all of which have slowed and complicated the recovery process.

We assisted local and national agencies and working groups in addressing the problems caused by landslides. The initial two weeks following the event were primarily focused on field surveys and rapidly assessment of the most critical situations that demanded immediate actions to ensure public safety. Subsequently, our efforts shifted towards landslide mapping. In a first stage, it was crucial to identify the roads and buildings affected by landslides to coordinate emergency interventions and perform an initial damage assessment. Afterward, we completed the landslide inventory to develop a comprehensive map detailing all landslides triggered by the event across the area. This map has been officially designated as the landslide map for the May 2023 event by the Po River Authority and the Emilia-Romagna region, and it is currently being used by the Commission for Reconstruction for implementing the recovery phase.

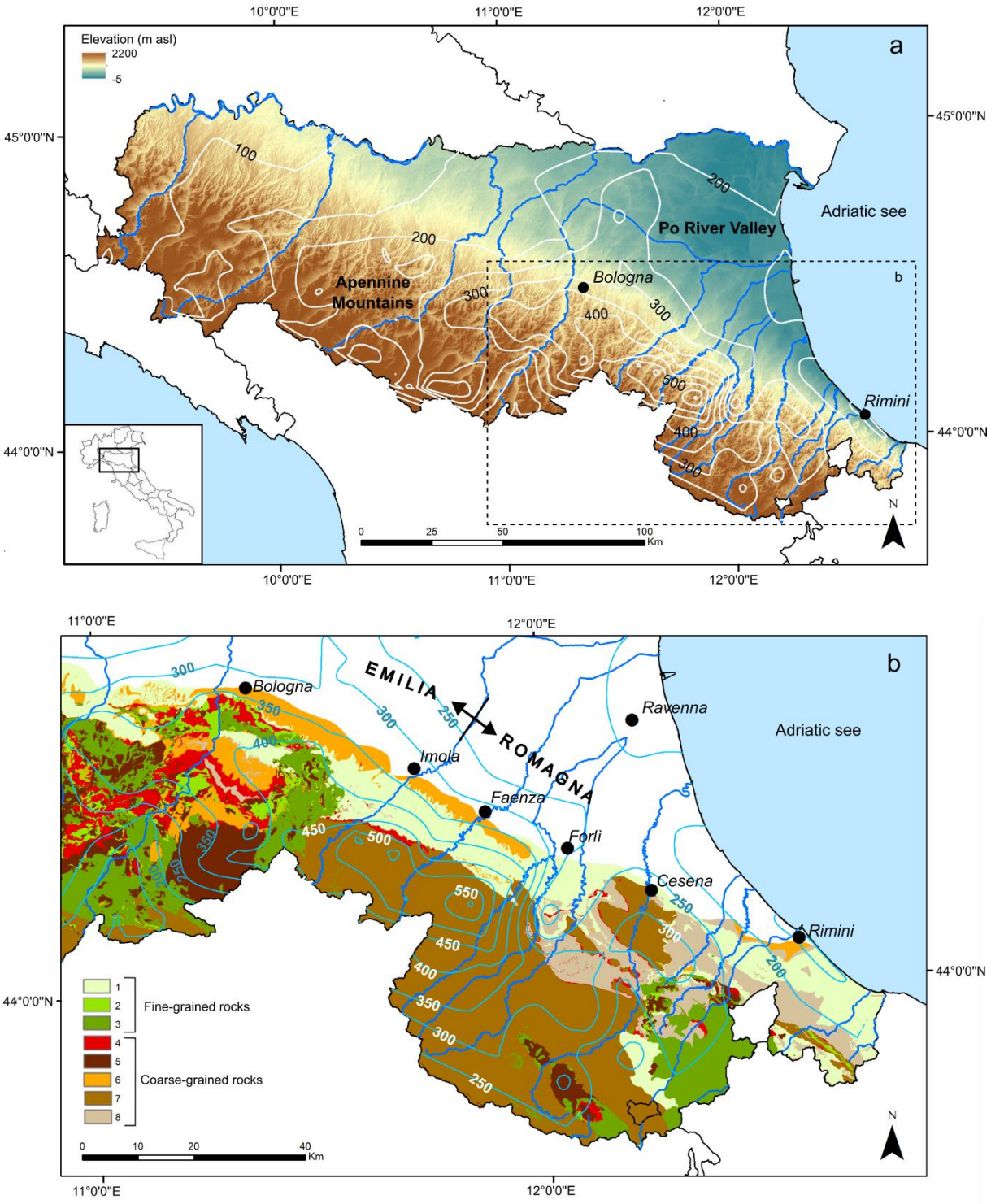


Fig. 3.1. **a**) Overview map of the Emilia-Romagna region (Italy), illustrating elevation and cumulative rainfall isohyets from May 1-17, 2023. **b**) Detailed view of the area most impacted by the event, featuring the geological units referenced in Table 3.1.

3.5 Methods

This section describes the methodology used to develop the landslide inventory for the May 2023 event. It covers the classification of lithological units, the identification and mapping of landslides, their classification, and the quality control and data harmonization procedures implemented.

3.5.1 Litho-technical units classification

In the study area, bedrock geology significantly influences the morphology of the slopes, the mechanical properties of the weathered soil layer, the vegetation cover and, consequently the proneness to slope instability. As a matter of fact, these factors played a crucial role in the behavior of the slopes during the May 2023 event, controlling the type and density of landslides. Consequently, bedrock geology is an essential base layer of our landslide inventory.

The geological map of the Emilia-Romagna region, created by the regional Geological Survey, includes more than 600 geological formations. These formations are distinguished by unique features that signify variations in depositional environments, composition, or geological age. The variety and detail of these formations illustrate the region's complex geological history and the precision employed in the map's creation.

For our inventory, we categorized all geological formations into eight distinct units, as depicted in Fig. 3.1b and detailed in Table 3.1. These units were delineated by merging lithological characteristics with their respective structural domains, recognizing that the same rock type can display varying structural and mechanical properties depending on its location within the orogenic sequence. For example, flysch rocks within the Ligurian domain (unit 5) are generally more fractured, less resistant, and prone to deep-seated landslides compared to those in the Tuscan-Umbrian domain (unit 7), due to the extensive tectonic stress they endured in the accretionary wedge.

The eight units identified were further divided into two broad categories: fine-grained rock masses (units 1 to 3) and coarse-grained rocks (units 4 to 8). This categorization aids in the preliminary differentiation of the types of weathered soil covers these rocks produce, which experienced widespread landslides during May 2023. Coarse-grained rocks typically produce granular soils composed of sand, gravel, and cobbles, with smaller amounts of silt and clay, aligning with the "debris" category in the Cruden and Varnes (1996) classification. In contrast, fine-grained rocks lead to the formation of fine soils predominantly made up of silt and clay, fitting the "earth" classification. These two categories, "debris" and "earth," are utilized to classify landslides that occurred on soil-covered slopes.

Table 3.1. Classification of the geological formations in the Emilia-Romagna region into eight units, based on their lithological composition and geological structural domains. Units 1 to 3 consist mainly of fine-grained rocks, while units 4 to 8 are primarily composed of coarse-grained rocks.

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

3.5.2 Landslide identification and mapping

Landslides identification and mapping was conducted manually using high-resolution aerial images. These images were captured using a Leica DMC III sensor aboard a Cessna 402C aircraft, flying at approximately 4700 meters above sea level. The images, taken shortly after the second rainfall on May 23, 2023, have a 0.2 m resolution and include four bands: RGB and near-infrared.

The mapping process was organized as follows. The total area was segmented based on the administrative boundaries of the municipalities. These sections were then distributed among three institutions: the University of Bologna, the University of Modena-Reggio Emilia, and the Geological Survey of the Emilia-Romagna Region. Each institution assigned four mappers, with a total of twelve individuals involved in the effort. Landslide detection was conducted in GIS environment by comparing pre- (April-July 2020) and post-event images with an on-screen zoom of approximately 1:1000. Once a landslide was spotted, further inspection was conducted using a 3D viewer with the high-resolution images overlaid on a 10 m DEM. Viewing the slope

from different angles enhanced the delineation of the affected area and the interpretation of the type of movement. Following this analysis, each landslide was classified into the specified classes described in the next section. The digital mapping of the landslide polygon was then executed at scales ranging from 1:800 to 1:200, depending on the landslide's size, ensuring precise tracing of the affected perimeter.

Identifying the landslides was relatively straightforward and objective, but delineating their boundaries was more subjective. Many landslides became fluidized upon failure, with the distal debris spreading among trees without removing vegetation, thus complicating the mapping of the deposit. Moreover, several slopes experienced complete removal of soil cover by adjoining shallow failures, blurring the distinction between individual slides. In these cases, we chose to interpret the landslide boundaries rather than just tracing the visible debris edges. For fluidized slides, the polygons were adjusted by connecting visible debris patches to include areas obscured by vegetation. For coalescent slides, we attempted to map each individual slide by identifying distinctive arcuate shapes along the detachment scarps that signified separate failures. Although this approach introduced some subjectivity into the manual mapping process, it was essential to create a dataset suitable for analyzing the morphometric features of the landslides.

The manual mapping process was demanding and labor-intensive. Initially, we focused on mapping landslides in the areas most severely impacted by the event, particularly around roads and urban centers. This priority was set to align with the Civil Protection Agency's needs to identify damages during the emergency response. This initial phase of mapping, which produced several "damage maps" for the affected municipalities, spanned the first two months following the disaster. Subsequently, the landslide inventory was expanded to cover the entire area over the following months. The complete process took approximately six months to finish.

3.5.3 Landslides classification

Right from the start of our work, identifying and classifying the landslides triggered by the event was recognized as a critical task. The primary challenge we faced was distinguishing between various types of debris slides and debris flows, since standard classification systems do not clearly separate them.

Figures 3.2 and 3.3 illustrate the problem. According to the Cruden and Varnes (1996) classification, all the six landslides depicted in the Fig. 3.3 can be classified as debris flows. Yet, clear differences are apparent between the upper (a1-3) and the lower three (b1-3). The latter are typical debris flows that start on a steep slope and stop as the slope decreases; the former, while starting similarly on steep slopes, demonstrate extensive propagation and complete fluidization of the deposit as they travel much further. None of these cases involve a well-defined channel, thus the classification proposed by Hungr et al. (2014) that distinguished

channelized debris flows from unchannelized debris avalanches does not apply here. A similar challenge presents with debris slides (Fig. 3.3). Current classifications fail to distinguish between slides of different degrees of mobility, a distinction that is clearly visible in the field. Some slides exhibited in fact high mobility, completely clearing the vegetation (upper pictures a1-3), while others show low mobility, indicated by minimal vegetation damage (lower pictures b1-3). Understanding the conditions that lead to these diverse behaviors is crucial for hazard assessment and necessitates differentiating these phenomena.

An additional classification challenge involves rock-block slides. These landslides impacted the homoclinal slopes of the Marnoso Arenacea Formation (unit 7 in Fig. 3.1b) and manifested as massive, translational rock-slab slides along bedding planes. While classifying these landslides poses no issues, it was necessary to distinguish between rock slides based on their degree of evolution. Some experienced movements ranging from several meters to tens of meters, signaling paroxysmal failures (Fig. 3.4 a1-a3), whereas others shifted merely a few centimeters, indicative of incipient, undeveloped failures (Fig. 3.4 b1-b3). The latter represent highly dangerous zones prone to potential collapse and thus required special attention.

These classification challenges were extensively discussed by our team. We ultimately decided to adopt the Cruden and Varnes (1996) classification system to define the primary types of landslides. These include debris slides (DS), debris flows (DF), and rock-block slides (RS) in the coarse-grained units (Fig. 3.1b), along with earth slides (ES) and earth flows (EF) in the fine-grained units. Then, we introduced the informal subclasses of: high-mobility debris slides (DS1), low-mobility debris slides (DS2), long-runout debris flows (DF1), limited-runout debris flows (DF2), fully-developed rock slides (RS1), and incipient rock slides (RS2) to capture the varied behaviors observed in the field. Subclasses were not assigned to earth slides (ES) and earth flows (EF) because landslides in areas with fine-grained soils were significantly less frequent and had milder impacts. This is because fine-grained soils have lower permeability and are more likely to fail during extended periods of rainfall rather than during brief events, where surface runoff and flooding are more prevalent.

Figure 3.2. Representative images of the two distinct types of debris flows caused by the May 2023 event.

Figure 3.3. Representative images of the two distinct types of debris slides caused by the May 2023 event.

Fig. 3.4. Representative images of the two distinct types of rock-block slides caused by the May 2023 event.

3.5.4 Quality control

To ensure the consistency of the results, an experienced geologist specialized in geomorphology and landslide mapping conducted a comprehensive review of the entire area after the completion of the manual mapping. This critical review focused on several key aspects to verify that all mappers adopted the same standards of detail and accuracy. Guided by a 1x1 km grid, the reviewer assessed: i) any missed landslides; ii) the precision in outlining landslide boundaries; iii) the consistency in interpreting vegetated areas; iv) the adherence to the established classification criteria; v) the accurate segmentation of individual slide events.

The findings from this review were summarized in a report sent to the twelve mappers. The report ranked the need for revisions in each municipality from "small" to "high" and included a detailed explanation of the necessary adjustments along with screenshots highlighting the errors detected. Each mapper was then tasked to revise their section of the manual inventory based on this feedback. This review and revision phase lasted approximately two months.

Following these adjustments, the manual landslide map was significantly improved. Although some variations persisted in the resolution of digitization and in interpreting boundaries obscured by vegetation, the primary discrepancies were effectively addressed and resolved. The

only remaining issue was the variation in landslide classification among mappers, that was evident when comparing the inventory maps of different municipalities. To address this, we developed the automatic procedure detailed in the next section.

3.5.5 Data harmonization

As mentioned earlier, the criteria for landslide classification were extensively discussed among us. A substantial number of cases were collectively examined and analyzed to synchronize the mappers' perceptions and foster a unified approach to class attribution. These preparatory efforts resulted in a substantial homogeneity in the classification of rock-block slide phenomena. However, notable variations were still evident in the final map between different types of debris slides and flows, which were influenced by the subjectivity of the mappers. Some determined the classification based on the landslide's shape, others on the presence of flow-like features in the deposition area, and yet others on the texture of the debris within the landslide polygon. On the other hand, such distinction is inherently subjective due to the gradual transition between slides and flows, making it challenging to establish a clear-cut boundary or to define strict classification rules. The same issue arose with classifying landslides based on their degree of mobility. While it was evident that many landslides exhibited complete fluidization and high mobility, significant discrepancies persisted in how the mappers categorized these events.

To address this issue, we implemented the automated procedure depicted in the flow chart of Fig. 3.5. This procedure utilizes standardized criteria to ensure uniform classification of landslides throughout the area and to correct inevitable errors in such a large dataset. The automated procedure was applied to all landslides except for rock-block slides, which have distinctive features that all mappers clearly and consistently recognized. Four key steps were identified to achieve a consistent classification of material type, movement type, degree of mobilization of debris slides and degree of fluidization of debris flows.

Material type (debris or earth)

The initial step was verifying the classification of material types. As previously mentioned, landslides on soil-covered slopes were divided into "debris" and "earth" categories according to the Cruden and Varnes (1996) classification. In our study area, this classification is clearly defined by the underlying bedrock geology; "debris" is derived from coarse-grained rock units, and "earth" from fine-grained units (Fig. 3.1b).

All mappers employed this classification system, referencing the geological map of the Emilia-Romagna region at a 1:10,000 scale, which provided an objective and standardized framework for classifying material types. However, manual mapping led to inconsistencies due to human

error and subjective judgments, particularly when categorizing landslides spanning multiple material types. To address these issues, we overlaid landslide polygons on the lithological map (Fig. 3.1b) and classified each landslide as either 'debris' or 'earth' based on its polygon's centroid location. This classification was achieved through a simple spatial join between the landslide data and the lithological map within the GIS environment.

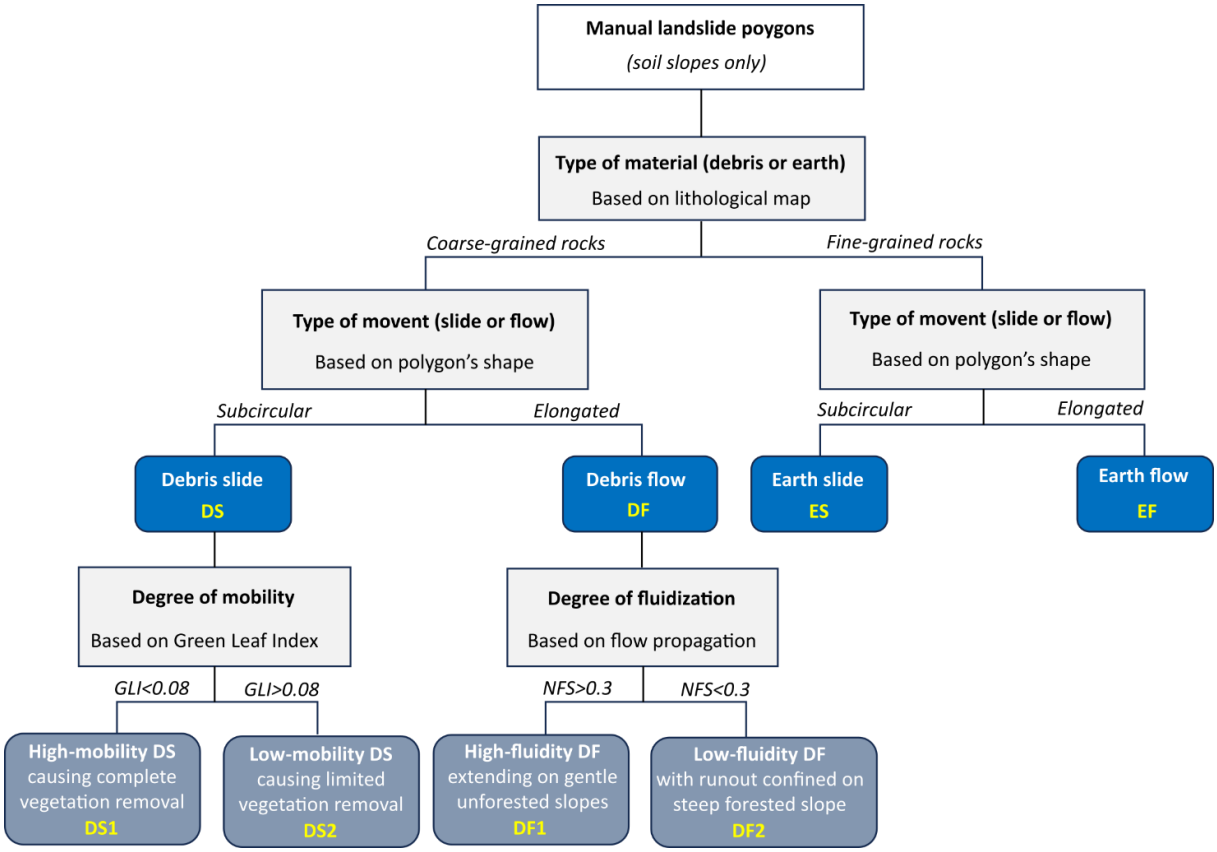


Figure 3.5. Flowchart depicting the process used to ensure data quality and standardize the classification of manually mapped landslides.

Type of movement (slide or flow)

To standardize the distinction between slides and flows, we employed a standard Convolutional Neural Network (CNN) specifically designed to recognizes the distinct shapes of slides and flows. The CNN was trained with data from the Casola Valsenio municipality. This area was chosen due to its highly accurate manual mapping and the thorough analyses carried out using multiple mapping techniques (Berti et al., 2024 - " Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy", under review). Moreover, Casola Valsenio served as the initial training ground for the mappers and is the area where classification challenges were collaboratively discussed.

The CNN features an input layer, two convolutional layers (each with batch normalization and ReLU activation), and subsequent max pooling layers to reduce image dimensions and enhance

feature extraction. The input layers processes 300x300 pixel black/white images of individual landslides, with landslide areas marked in white. These features are then categorized into 'slide' or 'flow' through a fully connected layer, followed by a softmax layer that determines the class probability. To enhance the model's ability to generalize, we implemented various data augmentation techniques, including random horizontal reflections, rotations ranging from -90 to 90 degrees, scaling from 80% to 120% of the original size, and translations up to 10 pixels.

The network was trained using a randomly selected half of the 4156 debris slides and 1115 debris flows identified in Casola Valsenio, while the other half was utilized to fine-tune the network's hyperparameters and to test and assess the model's performance. These evaluations showed that the CNN effectively replicates expert classifications of slides and flows. Utilizing the Adam optimizer with an initial learning rate of 0.001 over 100 epochs, the CNN reaches an F1-score of 0.80 on the testing dataset, indicating robust accuracy in terms of both precision and recall. Fig. 3.6 displays the confusion matrix obtained for the testing dataset, alongside a selection of landslide images that were correctly and incorrectly classified by the neural network. Of course, as clearly evident looking at the False Positives and False Negatives cases, the CNN cannot overcome the inherent ambiguity in classifying landslides that fall between slides and flows, particularly those that are only partially fluidized and whose polygon shapes are neither distinctly sub-circular nor clearly elongated. However, by implementing the network across all the polygons, we ensure that the classification criteria agreed in Casola Valsenio are consistently applied throughout the entire area.

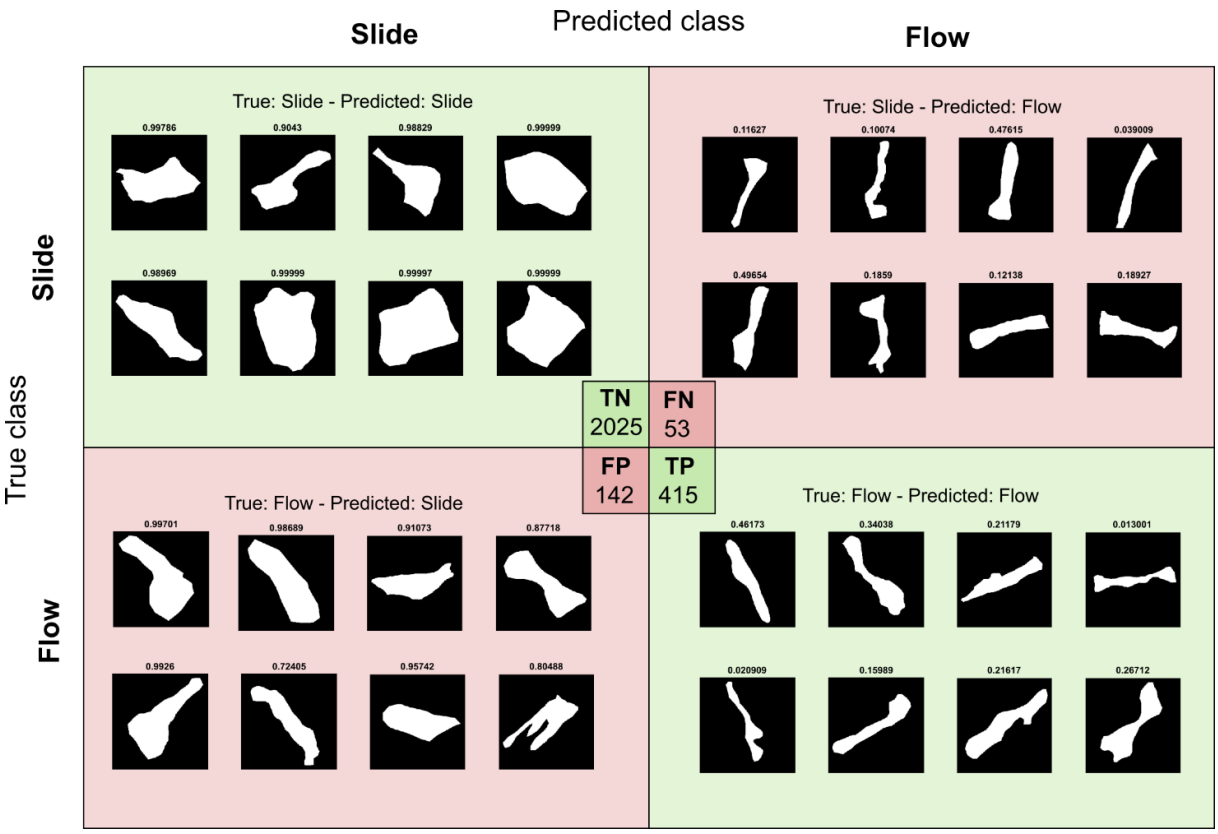


Figure 3.6. Outcomes from the Convolutional Neural Network model applied to differentiate slides from flows using the shape of the polygons. The figure displays the confusion matrix for the testing dataset, which includes 50% of the landslides manually mapped in the Casola Valsenio municipality: TN=True Negative; FN=False Negative; FP=False Positive; TP=True Positive. The small polygons in each category represent example landslides that are correctly (TP, TN) or incorrectly (FP, FN) classified.

Degree of mobility of debris slides

Debris slides were classified by mappers into two categories based on their apparent mobility (Fig. 3.3). The class DS1, indicating high-mobility slides, was assigned to slides that showed extensive internal disruption and complete removal of vegetation. The class DS2, denoting low-mobility slides, was assigned to slides with minimal internal deformation and little impact on vegetation. While mappers collectively agreed on these criteria for classifying debris slides, discrepancies arose due to variations in personal judgment.

To standardize this assessment, we evaluated the mobility of debris slides by examining the remaining vegetation cover after movement. The Green Leaf Index (GLI) was used to quantify the amount of green vegetation within a landslide polygon:

$$GLI = \frac{(2 \cdot Green - Red - Blue)}{(2 \cdot Green + Red + Blue)} \quad [4.1]$$

Here, Green, Red, Blue denote the reflectance values from the respective color bands. The GLI ranges between -1 and 1, with higher values indicating a denser presence of green leaves.

The choice of GLI over NDVI was primarily due to data availability. The NDVI requires a near-infrared (NIR) band, which was not consistently available across the entire study area. In contrast, GLI relies solely on the visible spectrum (RGB bands), ensuring uniform applicability to all datasets. This made GLI a more practical and consistent choice for our analysis, while still providing reliable information about vegetation cover.

To identify a suitable GLI threshold for distinguishing the two classes, we analyzed the frequency distribution of GLI values for DS1 and DS2 in the Casola Valsenio dataset. As shown in Figure 3.7, a distinct separation is observed in the higher categories: 99% of the 4144 high-mobility debris slides (DS1) have GLI values under 0.08, indicating they are primarily bare soils with minimal or no vegetation. In contrast, 34% of the 125 low-mobility slides (DS2) exceed this threshold, suggesting the presence of vegetation. Such a threshold therefore allows for an effective classification of DS1, but it risks misclassification of DS2.

The challenge in distinguishing the two classes stems from the inherent subjectivity involved, especially when vegetation is only partially removed. Like the difficulty in differentiating between slides and flows, no automated method can fully address this issue. However, in our case, the occurrence of DS2 is significantly less frequent than DS1. Consequently, we have chosen to set a GLI threshold of 0.08, acknowledging that this may lead to some misclassification errors with DS2. This approach classifies nearly intact vegetation slides as low-mobility (DS2) and those with partial vegetation as high-mobility (DS1). The resulting F1-score is 0.86, and this threshold has proven stable whether computed on a randomly selected subsample or a specific segment of the Casola Valsenio area.

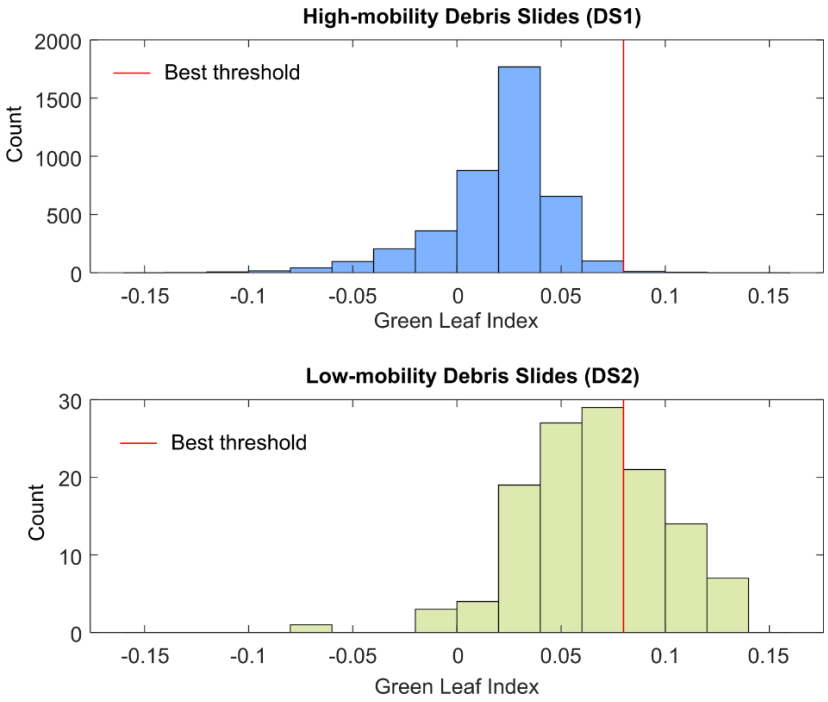


Figure 3.7. Comparison of the frequency distributions of the Green Leaf Index for high-mobility debris slides (upper) and low-mobility debris slides (lower) manually mapped in Casola Valsenio municipality. The red line marks the optimal threshold distinguishing the two landslide types.

Degree of fluidization of debris flows

Debris flows were divided into two distinct classes to highlight differences in fluidization and runout (Fig. 3.2). The DF1 category was used for long-runout debris flows, marked by fluidized deposits spreading over relatively flat terrain. Conversely, DF2 was used for debris flows with more limited fluidization, typically confined to steep, forested slopes. Mappers used these criteria but also looked at factors such as the size of the debris flow, the presence of a channel, or the location of the initiation area. As a result, the manual classification of DF1 and DF2 was notably inconsistent.

The classification problems are evident when attempting to define an automatic standardization procedure. Both DF1 and DF2 exhibit elongated shapes and absence of vegetation within the landslide areas. Consequently, previous methods that rely on polygon shape or vegetation cover are not applicable. One potential approach could involve using the mean slope of the landslide area, which is generally lower for DF1. However, this metric could introduce bias into the dataset, particularly when comparing the morphological characteristics of the different landslides.

After experimenting with various factors and machine learning techniques, we decided on a simple, reproducible method. These experiments included the evaluation of different input

layers and various training combinations, as well as the application of multiple machine learning models to identify optimal classification thresholds. While these efforts provided valuable insights, they are discussed in greater detail in Chapter 5, where their broader implications and applications are further explored. Using the Casola Valsenio dataset again, we determined that a reliable indicator of debris mobility is the percentage of the landslide area that extends over Non-Forested Slopes (NFS). NFS encompasses all slopes lacking forest cover, that in most cases are shrub and/or grassy areas, areas with sparse or no vegetation, and agricultural lands. Mappers typically classified debris flows that overrun these areas as DF1. NFS is simply given by:

$$NFS = \frac{A_{NF}}{A_{Tot}} \cdot 100 \qquad [4.2]$$

Here ANF is the landslide area on non-forested slopes and A is the total area the landslide. ANF was detected by overlapping the landslide polygon with the soil use coverage SU2014 provided by the Emilia-Romagna region. This coverage was derived from aerial images captured between May and September 2014, using four bands at a 0.5 m resolution, and classified according to the Corine Land Cover directive. All the slopes not categorized as 311 (Broad-leaved forest), 312 (Coniferous forest), or 313 (Mixed forest) were identified as non-forested.

In Casola Valsenio, a total of 1053 debris flows were documented. Among these, 471 were notably fluid and mobile (DF1), whereas the remaining 582 exhibited less mobility (DF2). The Non-Forested Slopes (NFS) values distinctly varied between the two classes, with DF1 generally displaying higher NFS values (Fig. 3.7). An NFS threshold of 0.3 has proven to be effective in distinguish between DF1 and DF2: 83% of the DF1 category exceed this threshold, whereas 82% of DF2 falls below it. The corresponding F1-score is 0.82, reflecting a high degree of accuracy.

The harmonization procedure described aboove resulted in significant modifications to the initial manual classifications. Approximately 50% of the debris slides with limited mobility (DS2) were reclassified as debris slides with high mobility (DS1) due to either heavy or partial clearing of vegetation cover by the movement. About 25% of debris flows (DF1 and DF2) were reclassified as debris slides (DS1 or DS2) due to the limited elongation of the deposit, and about 60% of earth flows (EF) were reclassified as earth slides (ES) for the same reason. It is important to stress that the harmonization process should not be viewed as an automatic classification but rather as an effort to apply consistent classification criteria across the entire area.

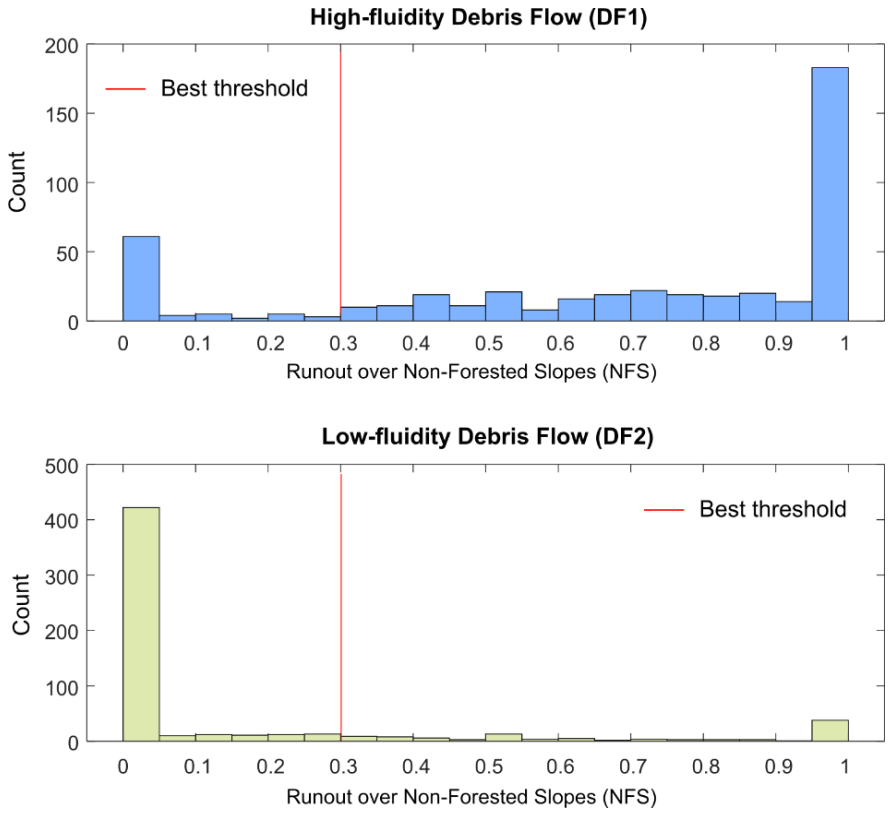


Figure 3.8. Comparison of debris flows with high fluidity (upper) and low fluidity (lower) in the Casola Valsenio municipality, analyzed through the ratio of runout over non-forested slopes (NFS). The red line indicates the optimal threshold for distinguishing between the two types of landslides.

3.6 The landslide inventory dataset

The landslide inventory for the 2023 Emilia-Romagna event is available as a shapefile in the Zenodo repository. The shapefile includes 80997 polygons, each categorized according to the classification described in section 3.5.2. The inventory encompasses landslides triggered by the combined rainfall events of May 1-3 and May 15-16, 2023, without distinguishing between the two events. Differentiation between the events is feasible only in specific small areas where high-resolution images were available after the first rainfall; however, this distinction is not included in the current dataset.

Figure 3.9 provides an overview of the inventory, showing landslide points (Fig. 3.9a) and a kernel density map (Fig. 3.9b). Notably there is a strong, though not perfect, correlation between cumulative rainfall and landslide density. In the eastern part of the region, known as Romagna, the 300 mm rainfall isohyet roughly outlines the area where landslide density exceeds 40 landslides per km². In contrast, the western part of the region, known as Emilia, has a landslide density below 40 landslides per km² despite receiving the same amount of rainfall. This difference can be attributed to the distinct geological settings of the two areas. As shown

in Fig. 3.1b, the Romagna region is primarily characterized by a Miocene flysch (Marnoso-Arenacea Formation, unit 7), which results in steep slopes and coarse-grained weathered soil. Meanwhile, the Emilia region has a more complex geological setting, including extensive areas of fine-grained rocks that responded less intensely to these rainfall events.

In the Romagna region, landslide density reached an impressive level of over 250 landslides per km². The zone most heavily affected, with more than 40 landslides per km², stretches across roughly 800 km² and covers the outer sector of the Marnoso-Arenacea Formation (red area in Fig. 3.10). About 64% of the landslides occurred within this zone. The landslide index, which is the ratio of landslide area to total area, reaches in this area the 20-25%. These figures are particularly significant considering they represent the percentage of the area destabilized during a single episode.

A deeper examination of the harmonized inventory underscores the occurrence and main features of various landslide types, providing insights into their spatial distribution and contributing factors. Figures 3.11 and 3.12 display several statistical details about the count, dimensions, and slope angles of these landslides. The results from these diagrams are discussed below, enhanced with additional observations from manual mapping and field surveys.

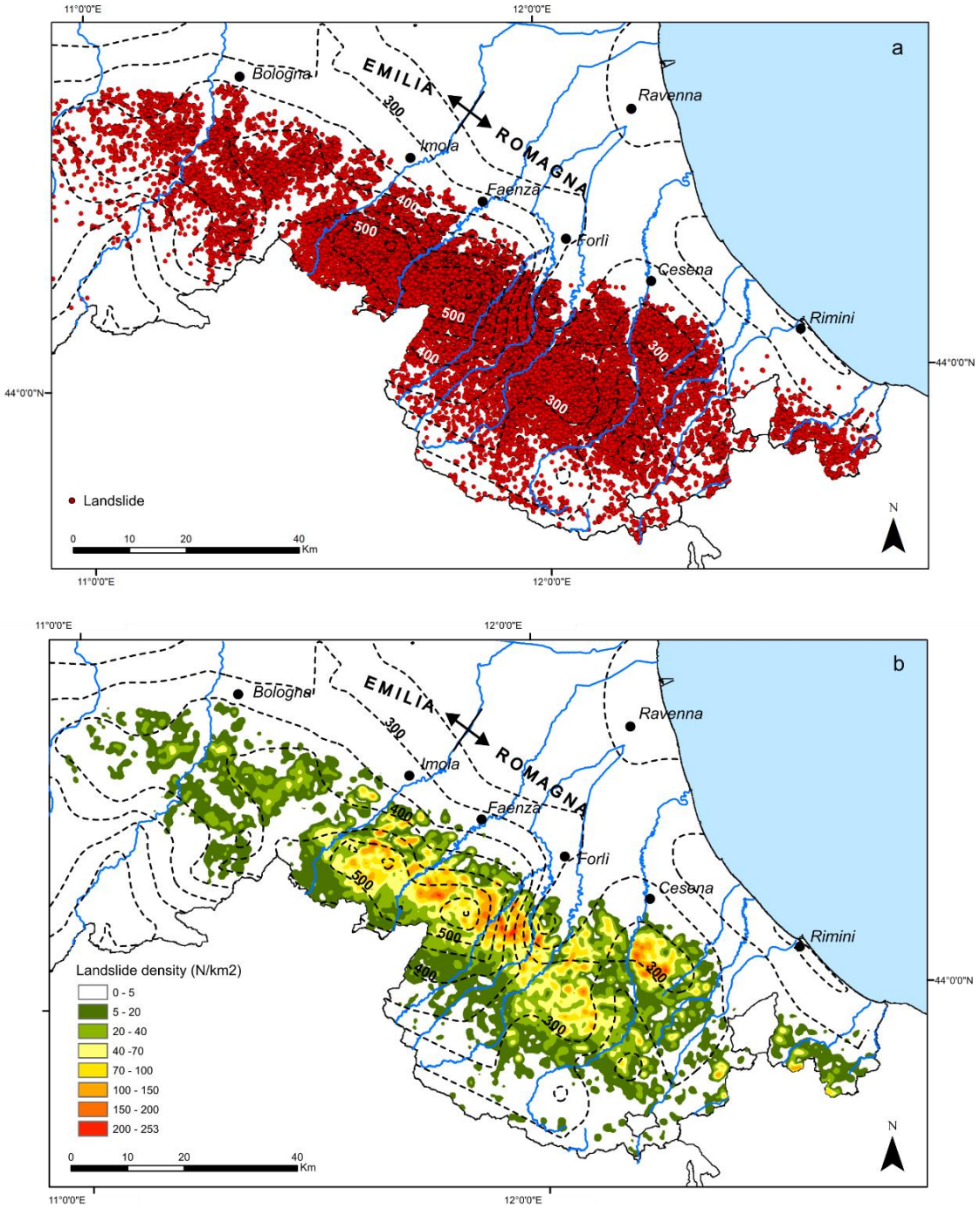


Figure 3.9. **a**) Map showing the distribution of the 80997 landslides triggered by the May 2023 event in the Emilia-Romagna region, manually mapped and represented as individual points. **b**) Density map calculated as the number of landslides per 1 km² cell.

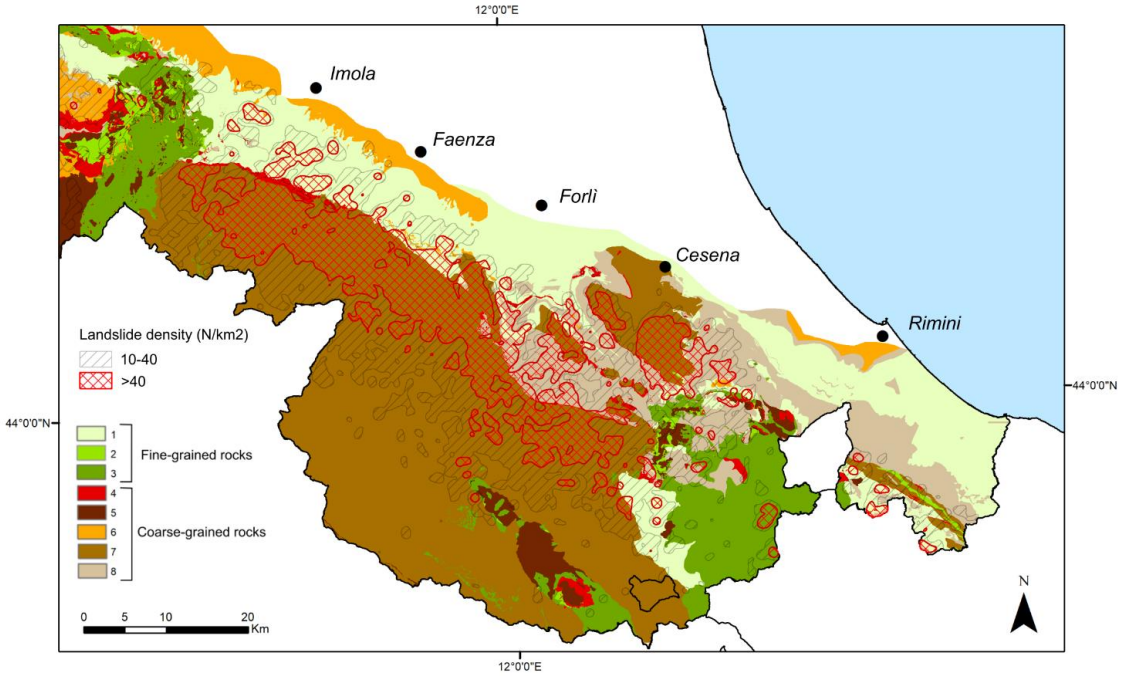


Figure 3.10. Detail of the Romagna area showing two classes of landslide density (between 10 and 40 landslide/km2 and more than 40 landslide/km^2) with the eight lithological units in background.

Debris slides (DS) represent 66% of all landslides by number and 49% by area, marking them as the most common type triggered by the event (Fig. 3.11). These landslides were generally small to very small (area less than 1000 m2, Fig. 3.12) and typically occurred on steep slopes with inclines exceeding 25°-30°. In the region, many of these slopes are covered with forests, as they are unsuitable for farming; hence, while root reinforcement and rainwater interception by the tree canopies exist, they were insufficient to prevent these failures. Approximately 94% of the slides were fast-moving and became liquefied after traveling a short distance (DS1). A minor fraction (6%) moved as a coherent mass, showing considerably less internal disruption (DS2). Slides with high mobility caused extensive damage to roads, buildings, and infrastructure and transported large amounts of debris and wood into rivers. Conversely, low-mobility slides predominantly occurred on milder slopes and near roadways. These slides might indicate early-stage slides that had not fully developed, secondary failures behind landslide headscarps, or slides involving rotational movements.

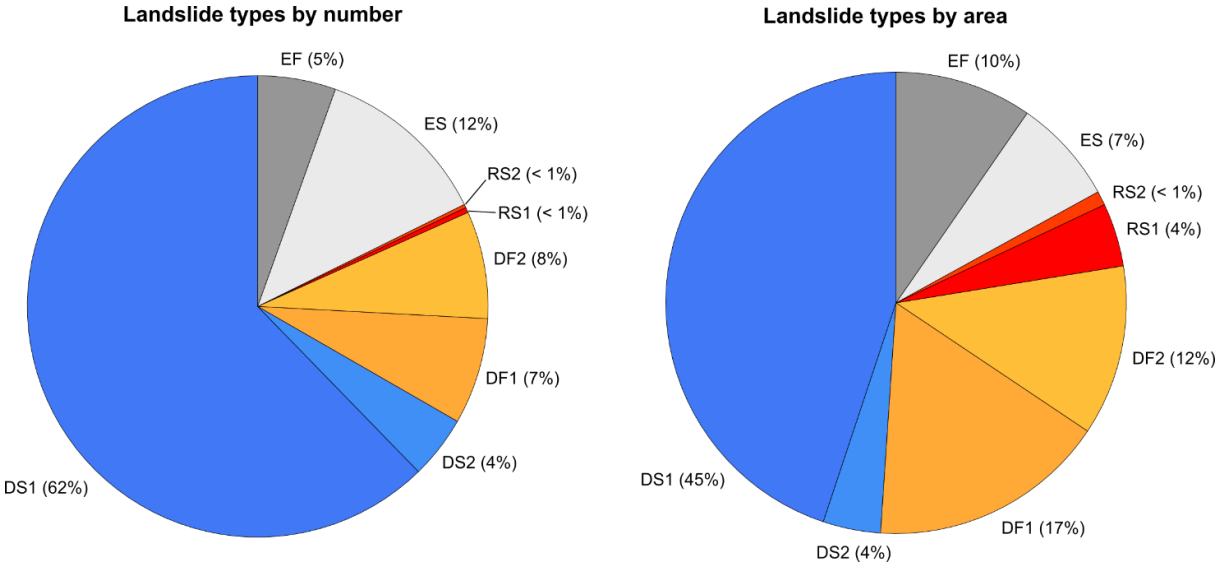


Figure 3.11. Pie charts showing the percentage of landslide types by number (left) and by area (right).

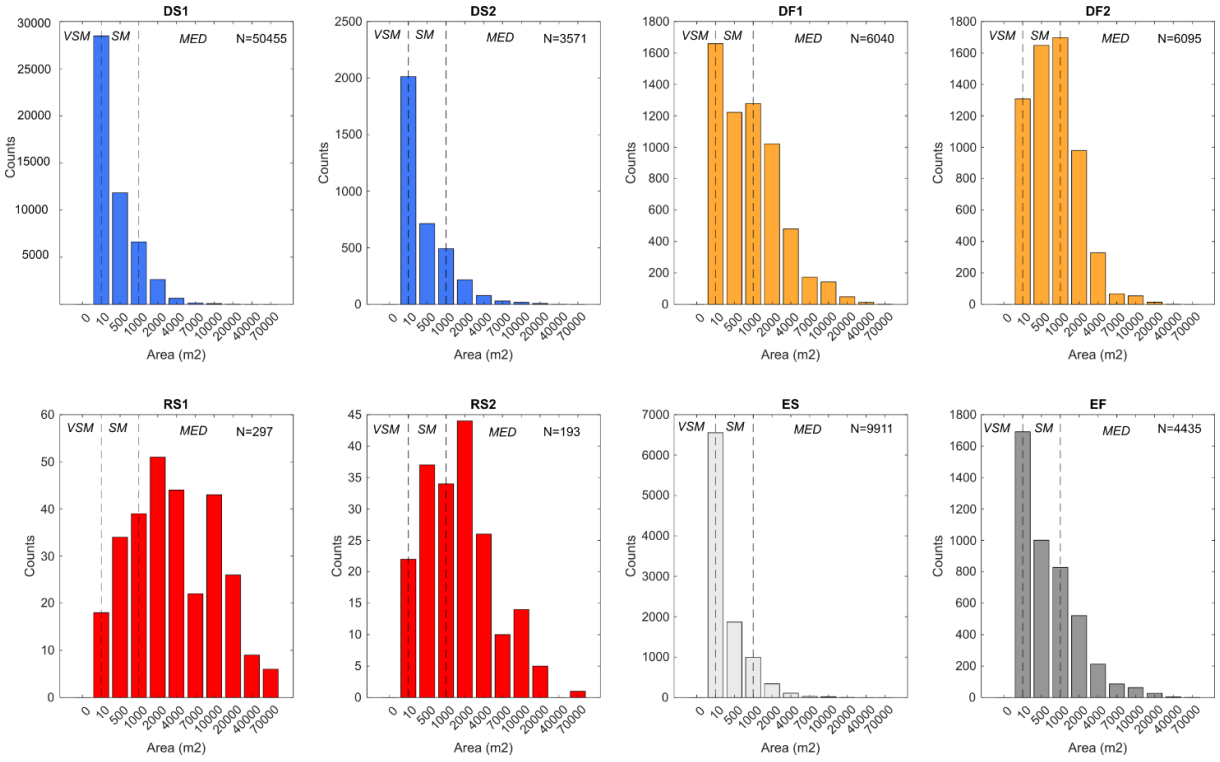


Figure 3.12. Frequency histograms depicting size of various landslide types. The classes labeled at the top of each chart (VSM=very small; SM=small; MED=medium) correspond to the size classification proposed by McColl and Cook (2024).

Debris flows (DF) are the second most frequent type of landslides, constituting 15% of the total count and 29% by area (Fig. 3.11). DF were consistently initiated by debris slides on slopes that are generally steeper than 25°-30°. Currently, it remains difficult to ascertain why some debris slides transformed into debris flows while others did not. However, it is evident that the predominant failure mechanism during the May 2023 event was shallow sliding of the weathered soil cover. Together, debris flows and slides represent 81% of the landslides cataloged in the inventory. After the initial failure, debris flows generally traveled without following predefined channels, and cleared the vegetation forming straight, elongated rectangular shapes. About 50% of these flows had relatively limited runouts halting along steep slopes (DF2), while the rest displayed significant higher mobility, spreading extensively across gentle slopes due to complete fluidization of the material (DF1). A notable feature of the debris flows triggered during the event, especially DF1, was their relatively low destructive power. In many cases, these flows approached buildings and roads without causing substantial damage and spread over grassy fields without harming the vegetation. The limited damage caused by these flows can be linked to their composition, primarily liquefied sand and silt without large cobbles or boulders. This composition enabled them to flow downslope as a dense slurry without a destructive bouldery front. The typical dimensions of debris flows range from "very small" to "medium" (Fig. 3.12).

Rock-block slides (RS) constitute less than 2% by number and 5% by area of all landslides (Fig. 3.11), but they left the most profound impression on the public and media. These landslides occurred on homoclinal slopes within the Marnoso-Arenacea Formation (lithological unit 7) and developed as planar slides along bedding planes that aligned with the slope. The thickness of the displaced rock mass varied from about 2 meters to over 30 meters, and several slides extended over areas larger than 10 hectares. Compared to debris flows and debris slides, rock-block slides affected more gentle slopes, typically less than 15° and were bigger in size (class "medium" Fig. 3.12). Their large volume, high velocity, and the fact that they occurred on sloping lands that were heavily urbanized and farmed made these landslides a major concern during the event. All rock slides initially traveled as coherent rock blocks, moving translationally for several to tens of meters. However, some slides disintegrated during their motion, transforming into rapid flows of debris and fragmented rock. This disintegration typically occurred when the displaced blocks tumbled down an existing scarp or struck a lateral slope, causing the material to break apart. These fragmented rock-block slides were highly mobile and covered long distances.

Earth slides and earth flows accounted for 12% and 5% of the total number of landslides, and 7% and 10% of the total area affected, respectively (Fig. 3.11). These landslides predominantly occurred within fine-grained units, specifically the Pliocene clays (unit 1) and Cretaceous clay shales (unit 3), as illustrated in Figure 3.1b. These regions generally experienced fewer landslides, with less severe impacts compared to areas dominated by coarse-grained rocks in

the southern parts. The distinct patterns of landslides in areas with coarse- and fine-grained lithological units are clearly illustrated in the sample maps of Figure 3.13. While landslides are commonly found on steep slopes in both cases, the coarse-grained units also show that even gentle slopes are impacted by extensive long-runout debris flows (DF1 in Fig. 3.13a) and rock-block slides (RS1). Conversely, in the fine-grained units (Fig. 3.13b), earth flows and earth slides are primarily concentrated in the badlands areas, with gentle slopes remaining largely unaffected. Moreover, in Emilia-Romagna, it is typical for earth slides and especially for earth flows to occur repeatedly at the same locations as reactivations of dormant landslides. This recurring pattern was evident during the May 2023 event, with most landslides appearing as reactivations of previously known landslides, which were already familiar to the local communities. In contrast, nearly all the landslides in the coarse-grained units—including debris slides, debris flows, and rock-block slides—represented first-time failures and occurred unexpectedly on slopes previously free of documented landslides.

3.7 Limitations and future updates

The landslide inventory was carried out with great care during the constraints imposed by the emergency situation and ongoing recovery efforts. Utilizing high-resolution 0.2 m imagery and an automated harmonization process has facilitated a detailed and consistent record across the region. However, as is typical with any expert-driven inventory, errors and inconsistencies are unavoidable.

Missed landslides, or false negatives, are likely to occur in shadowed areas like river gorges or steep slopes, as well as in forested areas where landslides have occurred without clearing the vegetation. Additionally, landslides with minimal ground displacement, although clearly visible on-site, may not be discernible in aerial images and thus may remain undetected. False positives—areas mistakenly identified as landslides—are also possible but expected to be fewer. These may include anthropogenic debris accumulations, excavation activities, or plowed fields that alter the soil surface in a manner similar to landslides, or landslides that happened after the pre-event images from April-July 2020 but before the May 2023 event. Nevertheless, we estimate that the combined total of missed or incorrectly identified landslides might constitute less than 1% of the total inventory.

The primary limitation of the inventory likely lies in the accuracy of the landslide boundaries. Not all mappers across the area had sufficient time to delineate the landslide polygons with high-resolution detail, resulting in some boundaries appearing jagged and imprecise upon closer inspection. The data quality procedures brought this issue to light, but redrawing all the rough-edged polygons would be excessively time-consuming. Given that the locations of these landslides are accurate, we chose to publish the inventory in its current state and defer any refinements to future versions.

Besides refining polygons, future updates will also incorporate changes recommended by local authorities, which are currently underway. In April 2024, the Emilia-Romagna region shared the landslide inventory with all municipalities affected by the event, requesting feedback on any overlooked landslides. This information is now being gathered, and a first update is scheduled for completion by the end of the year. Initial feedback primarily concerns small landslides that caused damage to private or public properties but were not detected in aerial photographs due to minimal displacement. These new landslides will be included in Version 2 of the inventory, which will be available in the same Zenodo repository.

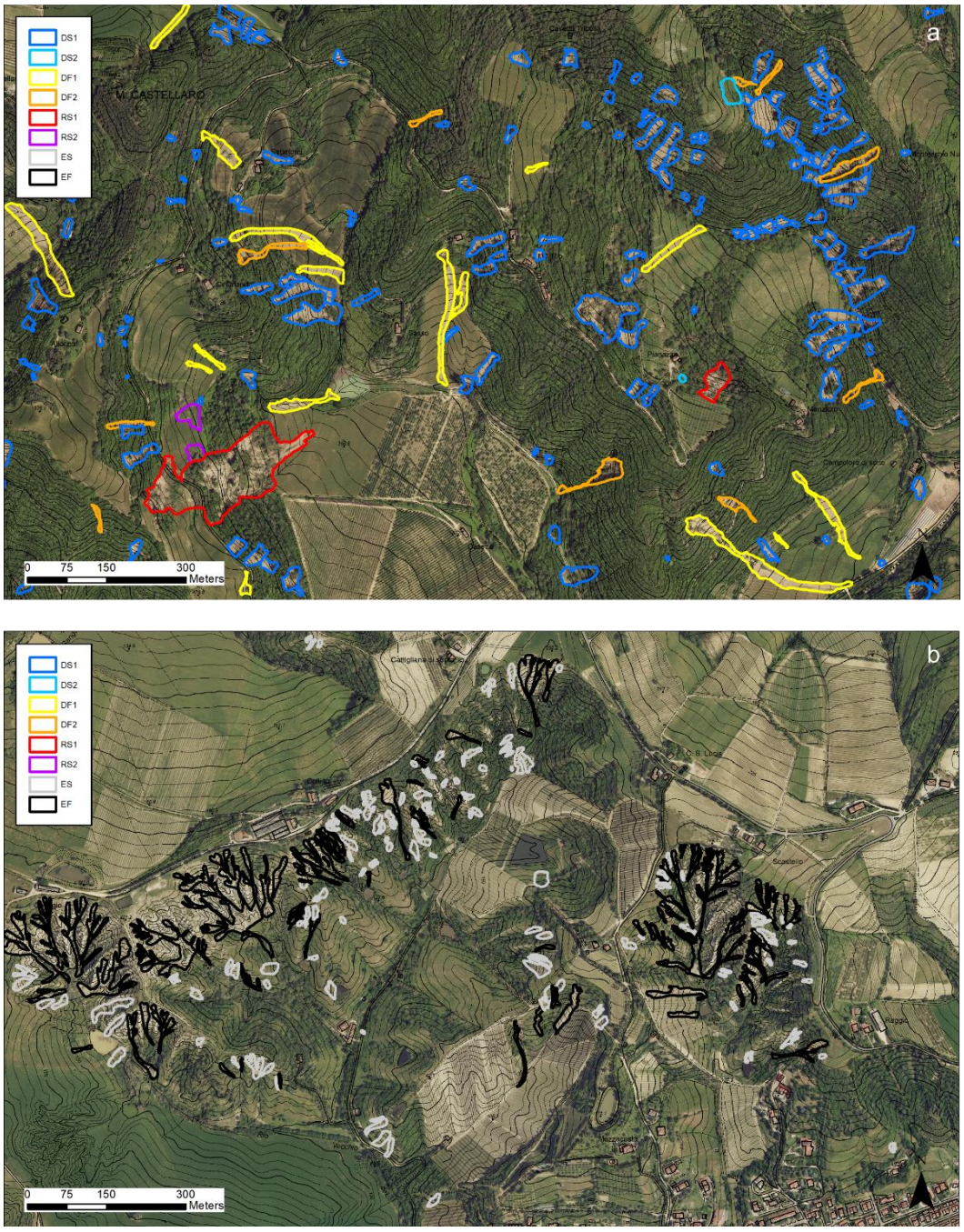


Figure 3.13. Example images from the landslide inventory showing two representative areas in coarse-grained units (**a**) and in fine-grained units (**b**).

3.8 Data availability

The landslide inventory is freely accessible in the Zenodo repository (DOI: 10.5281/zenodo.13742643, Pizziolo et al., 2024). The dataset is available as an ESRI (Environmental Systems Research Institute) shapefile and is compatible with GIS software. The shapefile encompasses several attributes: polygon ID (IDC), landslide type as manually classified by the operator (ClassMan), geological unit of the polygon's centroid (Lito), Green Leaf Index (GLI), percentage of deposit over Non-Forested Slopes (NFS), landslide type after applying the harmonization algorithm (ClassNew).

3.9 Concluding remarks

The dataset of the May 2023 Emilia-Romagna event encompasses more than 80.000 rainfall-induced landslides (mostly first-failure) distributed over an area of more 6000 km², with density reaching as high as 200 landslides/km2. Despite some inherent limitations and potential areas for improvement of the dataset, we believe that our landslide inventory offers significant value to the scientific community and to the involved institutions for several reasons.

Firstly, it documents the response of a large area to an exceptional meteorological event, likely linked to ongoing climate change. This can support the scientific community in proving that multiple occurrences of rainfall-related landslides are likely to become more frequent in the coming years, and can make decision makers more aware of the fact that even slopes that have been unaffected by landslides in the past, cannot be considered free of risk for the future.

Secondly, the Emilia-Romagna relatively straightforward geological framework makes it ideal for conducting geospatial analyses of landslide susceptibility, and to prove that they can be adopted to support land-use planning in addition to landslides inventories. Actually, the Emilia-Romagna region's geoportal provides free access to an extensive range of spatial data, including DEMs, lithology, land use, and rainfall data, all of which can be integrated with our landslide map to test both traditional and machine-learning-based predictive tools).

Thirdly, the predominance of shallow planar failures in this event provides an excellent case for testing physically-based slope stability models, and to highlight the relevance of such type of landslides in the study area, so to promote a much more careful evaluation of the possible impact of such phenomena on existing infrastructures network and for designing new assets.

In conclusion, we warmly invite interested colleagues to contact us with any questions, specific needs, or to initiate a collaborative research effort that could transform a tragic event into an opportunity to enhance our understanding of landslide risk assessment.

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).

3.10 References

Amatya, P., Kirschbaum, D., Stanley, T., Tanyas, H. (2021). Landslide mapping using object-based image analysis and open source tools. *Engineering Geology*, 282, 106000. https://doi.org/10.1016/j.enggeo.2021.106000

Ardizzone, F., Bucci, F., Cardinali, M., Fiorucci, F., Pisano, L., Santangelo, M., Zumpano, V. (2023). Geomorphological landslide inventory map of the Daunia Apennines, southern Italy. *Earth Syst. Sci. Data*, 15, 753-767. https://doi.org/10.5194/essd-15-753-2023

Auflič, M.J., Bezak, N., Šegina, E., Frantar, P. Gariano, S.L., Medved, A., Peternel, T. (2023). Climate change increases the number of landslides at the juncture of the Alpine, Pannonian and Mediterranean regions. *Sci Rep*, 13, 23085. https://doi.org/10.1038/s41598-023-50314-x

Bertolini G., Guida M., Pizziolo, M. (2005). Landslides in Emilia-Romagna region (Italy): strategies for hazard assessment and risk management. *Landslides*, 2, 302-312. https://doi.org/10.1007/s10346-005-0020-1

Bertolini, G., Corsini, A., Tellini, C. (2017). Fingerprints of Large-Scale Landslides in the Landscape of the Emilia Apennines. In: *Landscapes and Landforms of Italy*. World Geomorphological Landscapes, edited by Soldati, M. and Marchetti, M., 215-224, Springer, Cham. https://doi.org/10.1007/978-3-319-26194-2_18

Bhuyan, K., Tanyaş, H., Nava, L. Puliero, S., Meena, S.R., Floris, M., van Westen, C., Catani, F. (2023). Generating multi-temporal landslide inventories through a general deep transfer learning strategy using HR EO data. *Sci Rep*, 13, 162. https://doi.org/10.1038/s41598-022-27352-y

Bonham-Carter G.F. (1994). Geographic Information Systems for Geoscientists, Modelling with GIS. *Pergamon Press*, Oxford. ISBN 978-0-08-041867-4.

Brath, A., Casagli, N., Marani, M., Mercogliano, P., Motta, R. (2023). Rapporto della Commissione tecnico-scientifica istituita con deliberazione della Giunta Regionale n. 984/2023 e determinazione dirigenziale 14641/2023, al fine di analizzare gli eventi meteorologici estremi del mese di maggio 2023. Technical Report, Regione Emilia-Romagna, 147 pp. https://www.regione.emilia-romagna.it/alluvione/rapporto-della-commissione-tecnico-scientifica

Cardinali, M., Galli, M., Guzzetti, F., Ardizzone, F., Reichenbach, P., Bartoccini, P. (2006). Rainfall induced landslides in December 2004 in south-western Umbria, central Italy: types, extent, damage and risk assessment. *Natural Hazards and Earth System Sciences* 6, 237-260. https://doi.org/10.5194/nhess-6-237-2006

Catani F. (2021). Landslide detection by deep learning of non-nadiral and crowdsourced optical images. *Landslides*, 18, 1025-1044. https://doi.org/10.1007/s10346-020-01513-4

Ciccarese, G., Mulas, M., Alberoni, P.P., Truffelli, G., Corsini, A. (2020). Debris flows rainfall thresholds in the Apennines of Emilia-Romagna (Italy) derived by the analysis of recent severe rainstorms events and regional meteorological data. *Geomorphology*, 358, 107097. https://doi.org/10.1016/j.geomorph.2020.107097

Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.; Fotopoulou, S. (2014). Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; Pitilakis, K.; Winter, M.; Pastor, M.; Ferlisi, S.; Tofani, V.; Hervás, J.; Smith J.: Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ*, 73, 209-263. https://doi.org/10.1007/s10064-013-0538-8

Corsini, A., Ciccarese, G., Diena, M., Alberoni, P.P., Amorati, R. (2017). Debris flows in Val Parma and Val Baganza (Northern Apennines) during the 13 october 2014 alluvial event in Parma province (Italy). *Italian Journal of Engineering Geology and Environment*, Special issue V National Congress of AIGA (Cagliari, Italy), 29-38. https://doi.org/10.4408/IJEGE.2017-01.S-03

Crozier, M. (2005). Multiple-occurrence regional landslide events in New Zealand: hazard management issues. *Landslides*, 2, 247-256. https://doi.org/10.1007/s10346-005-0019-7

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

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102, 85-98. https://doi.org/10.1016/j.enggeo.2008.03.022

Ferrario, M.F., Livio, F. (2024). Rapid Mapping of Landslides Induced by Heavy Rainfall in the Emilia-Romagna (Italy) Region in May 2023. *Remote Sens*., 16, 122. https://doi.org/10.3390/rs16010122

Gaidzik, K., Ramírez-Herrera, M.T. (2021). The importance of input data on landslide susceptibility mapping. *Sci Rep*, 11, 19334. https://doi.org/10.1038/s41598-021-98830-y

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, 94, 268-289. https://doi.org/10.1016/j.geomorph.2006.09.023

Gariano, S.L., Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Sci. Rev*., 162, 227–252, https://doi.org/10.1016/j.earscirev.2016.08.011

Guzzetti, F., Cardinali, M., Reichenbach, P., Cipolla, F., Sebastiani, C., Galli, M., Salvati, P. (2004). Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. Engineering Geology 73, 229-245. https://doi.org/10.1016/j.enggeo.2004.01.006

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112, 42-66. https://doi.org/10.1016/j.earscirev.2012.02.001

Handwerger, A.L., Fielding, E.J., Sangha, S.S., Bekaert, D.P.S. (2022). Landslide Sensitivity and Response to Precipitation Changes in Wet and Dry Climates, *Geophys. Res. Lett*., 49, 1-12, https://doi.org/10.1029/2022GL099499

Hao, L., Rajaneesh, A., van Westen, C., Sajinkumar, K.S., Martha, T.R., Jaiswal, P., McAdoo, B.G. (2020). Constructing a complete landslide inventory dataset for the 2018 monsoon disaster in Kerala, India, for land use change analysis. *Earth Syst. Sci. Data*, 12, 2899-2918, https://doi.org/10.5194/essd-12-2899-2020

Hungr, O., Leroueil, S., Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, 11, 167–194. https://doi.org/10.1007/s10346-013-0436-y

Jaboyedoff, M., Oppikofer, T., Abellan, A., Pedrazzini, A. (2012). Use of LIDAR in landslide investigations: A review. *Natural Hazards*, 61, 5-28. https://doi.org/10.1007/s11069-010-9634-2

Luetzenburg, G., Svennevig, K., Bjørk, A.A., Keiding, M., Kroon, A. (2022). A national landslide inventory for Denmark. *Earth Syst. Sci. Data*, 14, 3157-3165. https://doi.org/10.5194/essd-14-3157-2022

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surf. Processes*, 29, 687-711. https://doi.org/10.1002/esp.1064

Notti, D., Cignetti, M., Godone, D., Cardone, D., Giordan, D. (2024). The unsuPervised shAllow laNdslide rapiD mApping: PANDA method applied to severe rainfalls in northeastern appenine (Italy). *International Journal of Applied Earth Observation and Geoinformation*, 129, 103806. https://doi.org/10.1016/j.jag.2024.103806

Piacentini, D, Troiani, F, Daniele, G, Pizziolo, M. (2018). Historical geospatial database for landslide analysis: the Catalogue of Landslide OCcurrences in the Emilia-Romagna Region (CLOCkER). *Landslides*, 15, 811–822. http://dx.doi.org/10.1007/s10346-018-0962-8

Rosi, A., Tofani, V., Tanteri, L., Tacconi Stefanelli, C., Agostini, A, Catani, F., Casagli, N. (2018). The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. *Landslides*, 15, 5-19. https://doi.org/10.1007/s10346-017-0861-4

Santangelo, M., Althuwaynee, O., Alvioli, M. et al. (2023). Inventory of landslides triggered by an extreme rainfall event in Marche-Umbria, Italy, on 15 September 2022. *Sci Data*, 10, 1–11. https://doi.org/10.1038/s41597-023-02336-3

Scorpio, V., Crema, S., Marra, F., Righini, M., Ciccarese, G., Borga, M., Cavalli, M., Corsini, A., Marchi, L., Surian, N., Comiti, F. (2018). Basin-scale analysis of the geomorphic effectiveness of flash floods: a study in the northern Apennines (Italy). *Sci Total Environ*, 640–641, 337–351. https://doi.org/10.1016/j.scitotenv.2018.05.252

Soeters, R., Van Westen, C.J. (1996). Slope instability recognition, analysis and zonation. Landslide Types and Processes, Transportation Research Board, *U.S. National Academy of Sciences*, Special Report, 247, 129-177

Steger, S., Brenning, A., Bell, R. et al. (2017). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides*, 14, 1767-1781. https://doi.org/10.1007/s10346-017-0820-0

Van Den Eeckhaut, M., Hervás, J., (2012). State of the art of national landslide databases in Europe and their potential for assessing susceptibility, hazard and risk. *Geomorphology* 139-140, 545–558. https://doi.org/10.1016/j.geomorph.2011.12.006

Zêzere, J. L. (2002). Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal), *Nat. Hazards Earth Syst. Sci*., 2, 73–82, https://doi.org/10.5194/nhess-2-73-2002

Zieher, T., Perzl, F., Rössel, M., Rutzinger, M., Meißl, G., Markart, G., Geitner, C. (2016). A multi-annual landslide inventory for the assessment of shallow landslide susceptibility – Two test cases in Vorarlberg, Austria, *Geomorphology*, 259, 40-54, https://doi.org/10.1016/j.geomorph.2016.02.008

Chapter 4

Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy

4.1 Preface

The preceding chapter explored the challenges in defining reliable rainfall thresholds for landslide prediction in regions with complex geological characteristics. This study highlighted the importance of integrating advanced data analysis methods with a deep understanding of the local geological context. The next step in this PhD project focuses on the practical application of this knowledge in the context of hydrogeological risk management.

This work required a phase of analysis of high-resolution remote sensing data, combined with field observations immediately following the catastrophic event of May 2023 in Emilia-Romagna. This chapter summarizes the analyses that led to the development and evaluation of automated methods for rapid landslide mapping in emergency situations.

The chapter is based on a paper currently under review on *Landslides* journal:

Berti M.[1], Pizziolo M.[2], Scaroni M.[2], Generali M.[2], Olivucci S.[3], Gozza G.[3], Formicola P.[3], Critelli V.[4], Mulas M.[4], Tondo M.[4], Lelli F.[4], Fabbiani C.[4], Ronchetti F.[4], Ciccarese G.[1], Dal Seno N.[1], Ioriatti E.[1], Rani R.[1], Zuccarini A.[1], Simonelli T.[5], Corsini A.[2]. Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy.

[1] Department of Biological, Geological, and Environmental Sciences, University of Bologna, Bologna, Italy.

[2] Regione Emilia-Romagna, Area Geologia, Suoli e Sismica, Bologna, Italy.

[3] Regione Emilia-Romagna, Settore Innovazione Digitale, Dati, Tecnologia e Polo Archivistico, Bologna, Italy.

[4] Department of Chemical and Geological Sciences, University of Modena and Reggio-Emilia, Modena, Italy.

[5] Autorità di Bacino Distrettuale del Fiume Po, Parma, Italy.

4.2 Abstract

Rapid and accurate mapping of landslides is essential for prompt response to emergencies triggered by rainfall or seismic events. This task can be done manually by an expert team, yet in scenarios involving large disaster areas it demands substantial time, may face difficulties in identifying minor landslides, and involves subjective judgement. Automated mapping is a promising alternative, offering the potential for quicker and more efficient map production, though concerns about its accuracy persist. This research examines the impact of the May 2023 cyclonic events in Italy's Emilia-Romagna region, which caused significant landslides and flooding. It evaluates the effectiveness of automated mapping algorithms, specifically the traditional NDVI change and the advanced U-Net neural network, over an 84 km² area where over 5200 landslides occurred due to the disaster. Both methods were implemented by considering the usual limitations of emergency situations, emphasizing the critical nature of time and the potential scarcity of a comprehensive ground-truth dataset for model training. The results show that U-Net, utilizing high-resolution images, outperforms NDVI in terms of accuracy, achieving an F1-score greater than 0.6. This enhanced performance of U-Net is ascribed to its proficiency in detecting landslides not only by changes in vegetation but also by analyzing the landslide's physical shape. However, the research ultimately determines that in the case of the May 2023 disaster, automated mapping processes cannot eliminate the need for manual verification, especially due to the critical necessity for precise damage identifications.

4.3 Introduction

As for any given task, doing it timely and well is not exactly easy, and rapid and accurate landslide mapping over large areas is no exception to this rule. This is especially true if a large number of landslides must be mapped to support emergency management after regional-scale rainfall or earthquake events. In such cases, remote sensing can improve rapidity in mapping (Guzzetti et al. 2012; Behling et al. 2014; Tanoli et al. 2017; Gorelick et al. 2017; Ghorbanzadeh et al. 2020; Ferrario et al. 2024; Novellino et al. 2024). However, upmost care should be given to minimize the number of missed landslides as well as of false landslides, so not to undermine the utility of such maps. The ability to quickly and accurately identify landslides depends on several key factors, including: (1) immediate access to high-resolution, cloud-free satellite or airborne images; (2) the deployment of a skilled team capable of visually analyzing these images to establish actual conditions on the ground; (3) the effectiveness of image analysis software in automatically detecting and outlining landslides; (4) the challenge in identifying some landslides through manual or automated methods using satellite or airborne images, due to their small size, obscuration by tree cover, or minimal impact on the terrain and landcover, making them undetectable without on-site inspections.

In Europe, the Copernicus Emergency Management Service (CEMS) makes use mid- to high-resolution pre- and post-event satellite imagery (such as Sentinel2 MSI at 10 m resolution in visible and NIR and Pléiades-1A/B at 0.5 m resolution) in order to provide Rapid Mapping of landslides and flooded areas using both visual and automatic methods (Joubert-Boitat et al. 2020). Unfortunately, only a limited portion of the imagery Copernicus uses is accessible to third parties (such as Civil Protection Authorities or their scientific collaborators) for further examination or for running automated landslide detection algorithms. For instance, datasets from Sentinel 2 are freely distributed by the European Space Agency, ensuring their quick availability (Shahabi et al. 2022; Satriano et al. 2023). Instead, acquiring higher resolution datasets, like those from Pléiades 1A/B, requires specific orders and faces distribution restrictions by Copernicus due to commercial contracts with external satellite data suppliers. This limitation significantly complicates the timely access to these datasets.

Similarly, in the United States, services of Rapid Response Landslide Inventory are provided by the USGS (Martinez et al., 2021) using visual interpretation of mid- to high-resolution pre- and post-event satellite imagery including Sentinel2, WorldView (0.3–0.5-m resolution), Planet (2.7–4.0-m). In Great Britain, Williams et al (2018) performed a satellite-based emergency mapping of landslides triggered by the 2015 Nepal earthquakes using visual interpretation of a wide range of mid- to high-resolution satellite imagery. They argued that highest-resolution data may not always be the most appropriate for wide-area mapping of landslides and that the time taken to produce outputs from their mapping campaign was most influenced by cloud-free image availability over the area of interest.

Automated landslide recognition algorithms based on image analysis are a possible alternative to visual interpretation and manual digitization of landslides to increase mapping rapidity over large areas (Guzzetti et al. 2012; Behling and Roessner 2017; Jain et al. 2024). For instance, setting thresholds in the changes of the NDVI (Normalized Difference Vegetation Index; Yang and Chen 2010; Niyokwiringirwa et al. 2024) from pre- to post-event imagery, is a well-established approach for semi-automated mapping of landslides that have significantly altered land-cover conditions (Mondini et al. 2011; Notti et al., 2023). Other prevalent techniques include pixel-scale unsupervised classification (Li et al., 2016; Zhong et al., 2020), Object-Oriented Analysis (Wu et al., 2014; Dou et al., 2015; Keyport et al., 2018), and methods that integrate various approaches (Adriano et al., 2020; Amatya et al., 2021). Recently, the development of high-performing image-segmentation and classification algorithms based upon Machine Learning (ML) and Deep Learning, (DL) has boosted the production of research papers aimed to demonstrate the possibility of automated landslide mapping over large areas (Prakash et al. 2020; Su et al. 2021; Prakash and Manconi 2021), even by using open-source resources such as U-Net, a convolutional neural network (NN) developed for biomedical image segmentation (Ronneberger et al., 2015). These methods always rely on a dataset of ground

truths (i.e. manually mapped landslides) to calibrate NDVI changes thresholds or to train ML, DL or NN algorithms.

In May 2023, the Emilia-Romagna Region in Italy (Fig. 4.1a) faced one of its most critical hydrogeological crises due to two cyclonic events of high intensity. These occurrences on May 1st-3rd and May 16th-17th led to rainfall reaching up to 500 mm on the Romagna Apennines, a significant increase compared to the area's average annual rainfall of approximately 1200 mm. The intense rainfall resulted in approximately 80,000 landslides across the slopes and widespread flooding in the cities of Imola, Forlì, Cesena, Ravenna, and vast portions of Romagna's floodplain sector. The aftermath included damages worth several billion Euros and 17 fatalities. Our involvement in the crisis began from the outset, assisting the Regional Agency for Civil Protection. A significant aspect of our efforts was dedicated to the surveying and mapping of the numerous landslides triggered by these events. We performed landslide mapping manually, utilizing ultra-high resolution images with a 0.2 m pixel resolution and a 3D viewer to accurately determine the boundaries and types of landslides. This meticulous effort resulted in a detailed dataset of all the landslides that occurred throughout the Romagna area due to the event.

This study aims to capitalize such ground-truth dataset to assess the potential benefits and drawbacks of using automated algorithms for mapping purposes, which should provide an efficient alternative to the laborious and slow process of manual mapping. Specifically, the paper explores the effectiveness of NDVI-change and U-Net methods in emergency scenarios, where the rapid and accurate detection of landslides is critical. In such situations, merely having a map of "potential landslide locations" is insufficient due to the high precision required for identifying damaged infrastructure, evacuating hazardous areas, and coordinating emergency responses. Automated mapping techniques are therefore expected to deliver robust results quickly without the need for extensive calibration. This application significantly differs from the typical focus in the literature, which is on developing new methods or refining existing ones under non-emergency conditions, free from the urgent time constraints and critical issues faced during emergencies. Our research aims to bridge the gap between theoretical studies and real-world applications of automated landslide mapping, or at least to spark a discussion on this topic within our community. The study specifically examines several crucial factors observed during the May 2023 disaster: i) the need for minimal calibration data in automated methods; ii) the impact of image resolution and quality; iii) the ability of automated techniques to detect various types of landslides; iv) the use of these methods for identifying damages caused by landslides. Assessing these elements is essential to determine whether automated approaches can substantially improve upon traditional, manual mapping methods, particularly in managing widespread landslide emergencies effectively.

4.4 The Romagna 2023 disaster

4.4.1 Triggering meteorological events

A detailed description of the triggering meteorological events that have affected the Emilia-Romagna Region during May 2023 has recently been published by Foraci et al. (2023) and in the technical reports referenced therein. Essentially, two primary events took place from May 1st to 3rd and from May 16th to 17th, with intermittent rain showers affecting the same areas on multiple instances in the intervening period. The event from 1-3 May saw intense rainfall targeting the central-eastern hilly areas, where cumulative rainfall approached 200 mm over 48 hours. The peak rainfall during this period was recorded at 274 mm in the Forlì mountains (Fig. 4.1b). Over half of the rain gauges in this central-eastern hilly sector registered two-day cumulative rainfall figures that exceeded historical maximums, with the rainfall intensity over a 24-hour period surpassing the estimated return period of 100 years. The event on May 16-17 impacted an even broader area of Emilia Romagna, with rainfall ranging from moderate to heavy intensity. Over 48 hours, cumulative rainfall exceeded the historical highs recorded at the start of May, with peaks between 212 mm and 261 mm in the mountains and hills of Ravenna, Forlì, and Cesena (Fig. 4.1c). In the intervals between these two significant events, additional rainfall episodes occurred across the region, notably between May 9-10 and May 12-14.

Overall, the total rainfall recorded from May 1 to May 17 amounted to a regional average of 221 mm, marking the highest 17-day cumulative rainfall in Emilia-Romagna since 1961 when compared to the region's climatological records. In the Romagna area, 80% of the rain gauges set new historical rainfall records, with average amounts exceeding 300-400 mm over the 17-day span and peaking at 610 mm.

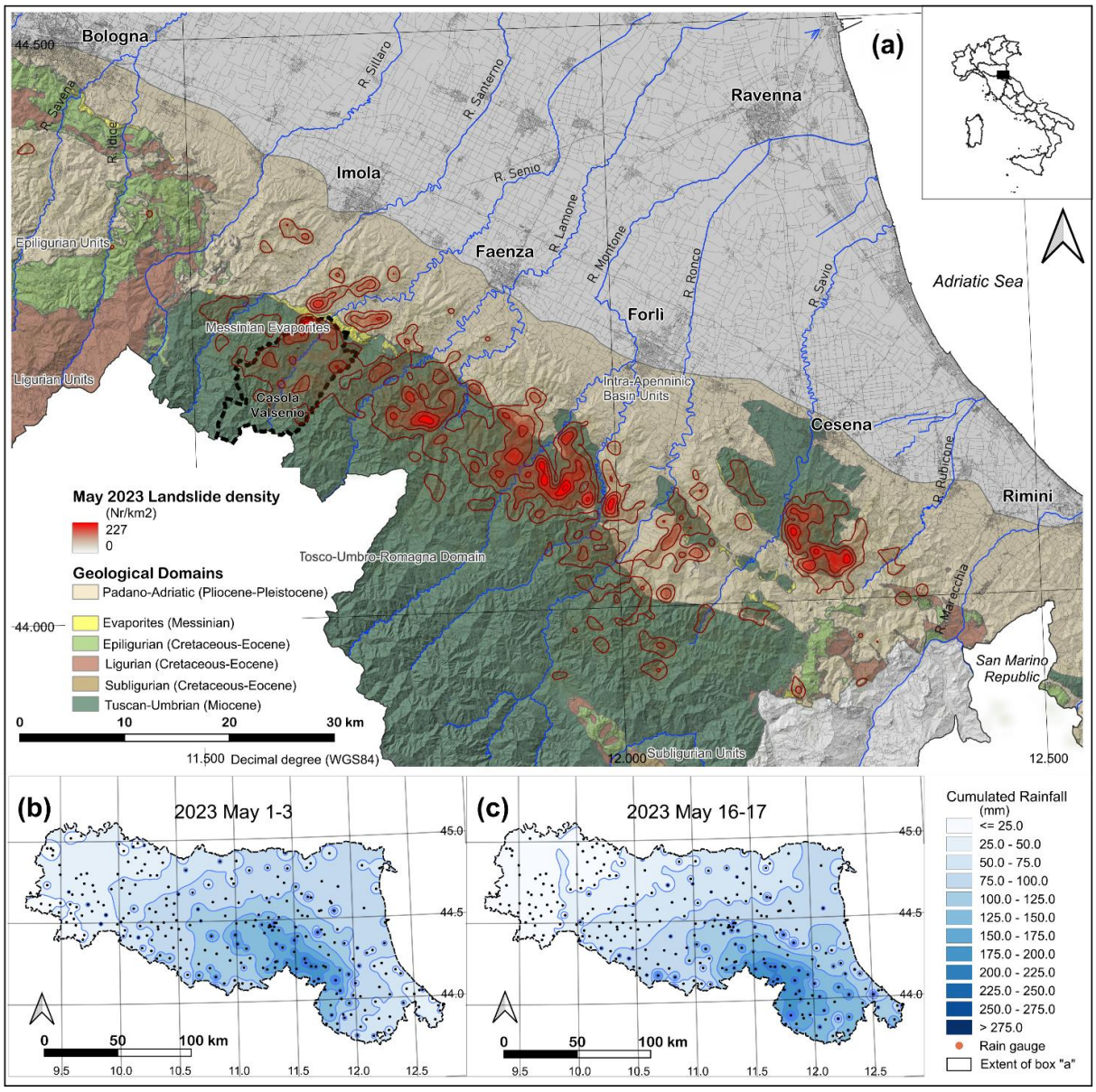


Figure 4.1. Overview of the May 2023 disaster in the Emilia-Romagna region. **A**) Distribution of landslide density determined through manual mapping; **b**) cumulative precipitation map for the first rainfall event; **c**) cumulative precipitation map for the second.

4.4.2 Effects on slopes stability

The meteorological events of May 2023 had widespread and profound effects on slope stability, with the geological composition of the area playing a crucial role. Specifically, the Romagna area is distinguished by the extensive occurrence of the Marnoso-Arenacea Formation (FMA; Burdigalian-Tortonian, Miocene), as illustrated in Fig. 4.1a. This formation primarily consists of an arenaceous flysch, composed by layers of arenite that range from centimeters to meters in thickness (A), interspersed with layers of pelite (P) (Ricci Lucchi, Valmori 1980). The ratio of arenite to pelites layers (A/P) varies, ranging from greater than 1 to over 10.

In the area, the Marnoso-Arenacea Formation is primarily characterized by a monocline structure with a dip of approximately 10°-15° towards the North-East, covered by a thin layer of weathered sandy to silty debris (Fig. 4.2). This geological configuration creates distinctively asymmetrical slopes, with mild slopes where the bedding planes run parallel to the slope's surface, and steeper inclines where the strata dip into the slope (Fig. 4.2). During the event, dip slopes mainly experienced large-scale translational rock-block slides (RS), with depths exceeding 10 meters. In contrast, steep slopes with strata dipping into the slope predominantly saw a high occurrence of debris slides (DS) characterised by high-mobility. These slides, typically shallow, extensively removed vegetation across broad areas exposing the underlying bedrock. Many debris slides evolved into debris flows (DF), traveling several hundreds of meters down, either confined within existing gullies or steep forested slopes or spreading unconfined over the slopes. In this latter case, the morphological characteristics closely mirror those of debris avalanches reported in literature (Hungr et al., 2013). However, unlike the cases documented in these studies, the scouring and embedding of material along the propagation path was in most of our case limited.

It is important to note that rock slides, debris slides, and debris flows, which accounted for the vast majority of the landslides in May 2023, were nearly all first-time failures (Pizziolo et al., 2023). Even in instances where these landslides overlapped with previously mapped landslides, it was almost exclusively because the new landslides expanded across areas of pre-existing landslide deposits. An exception was observed with earth and mud flows (EF) occurring in the clay formations exposed in the northern sector of the Romagna Apennines (Fig. 4.1a) and in the far western regions. Here, the May 2023 landslides often represented either partial reactivations or further developments of previously dormant or active landslides.

The landslides of May 2023 caused significant changes in the landscaper, exposing vast portions of slopes by removing the vegetation and soil cover. Many of these landslides featured rapid movement of the dislodged debris or rock masses. As a result, more than 1000 roads throughout the Romagna region were affected, with impacts occurring both in the initial detachment zones, along the flow track, or in the deposition zone. Due to their large volume and high speed, certain rock-block slides resulted in the total destruction of houses and led to fatalities. The sudden and rapid nature of these events created a widespread state of emergency, compelling many individuals to evacuate their homes and hampering access to villages or valleys due to the extensive damage to roads and infrastructure.

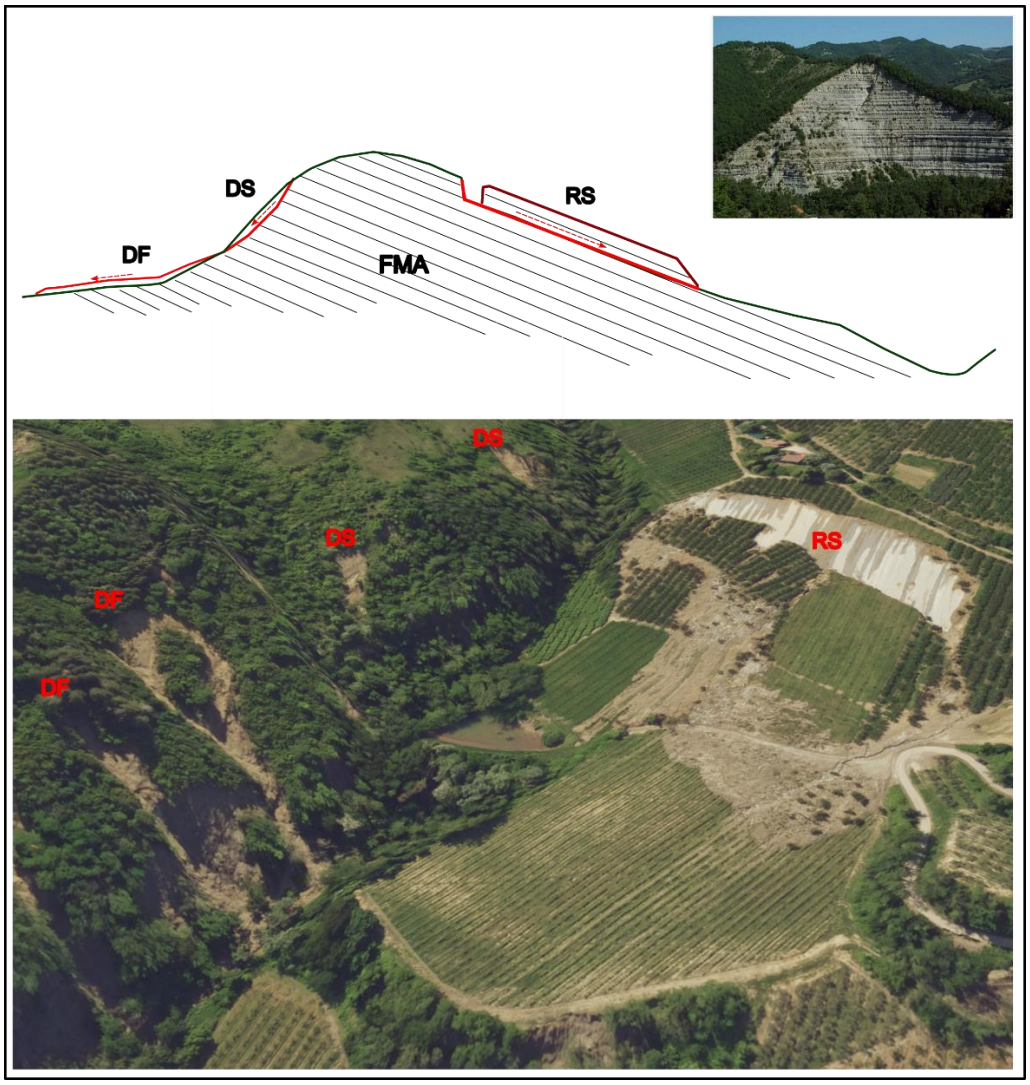


Figure 4.2. Representative landslides from the May 2023 Event. The top illustration provides a conceptual cross-sectional view of a slope within the Marnoso-Arenacea Formation (FMA), showing typical landslides occurring on flanks where bedding planes dip into the slope (left) and where bedding planes are parallel to the slope (right). Pictures illustrate various examples of landslide types. DS=debris slide; DF=debris flow; RS=rock-block slide. The upper-right picture shows the typical aspect of the Marnoso-Arenacea Formation exposed on a subvertical cliff.

4.4.3 Activities during the emergency

Immediately after the events of May 1-3 and 16-17, and during the subsequent recovery phase, the creation of a landslides inventory and the identification of damages proved to be a vital yet challenging task in emergency response. The Civil Protection Agency of the Emilia-Romagna region tasked us with creating a comprehensive inventory of the landslides caused by the event to coordinate emergency interventions and assess the damage to roads and buildings. The creation of the landslide inventory began immediately after the events, utilizing "landslide

notices" from the affected municipalities and media reports, alongside field surveys conducted by public institutions, universities, and professionals. However, given the extremely high number of landslides, it quickly became apparent that completing this task would require the interpretation of satellite or aerial imagery.

Soon after the 1st-3rd May event, Rapid Mapping of the Copernicus Emergency Management Service was activated for seven Areas of Interest (AoI), including three within the Apennines (EMSR659 - Flood in Emilia-Romagna, Italy). However, the landslide maps produced, which were of the "grading level" detail using post-event Pléiades-1A/B imagery at 0.5 m resolution, did not achieve the necessary completeness and accuracy for civil protection efforts. Notably, several known landslides were not included (false negatives), and areas with vegetation changes from other causes were marked as potential landslides (false positives). Additionally, after the second rainfall event, the National Civil Protection issued a map of "possible landslides" using pre- and post-event Sentinel 2 data (10 m pixel size) and NDVI-change analysis, following the method by Notti et al. (2023). The map was necessarily based on thresholds not calibrated on ground truths of the specific event, resulting in significant omissions and inaccuracies.

To create an accurate landslide map, a decision was made to compile a comprehensive inventory of landslides and associated damages through the analysis of satellite and aerial images (both visible and NIR spectra) by an expert team. This team carried out field surveys in the immediate aftermath of the event, thereby gaining direct knowledge of the landslide types to identify within the imagery. Initially, Sentinel2 data (with a 10 m resolution covering the entire affected region) and Planet data (with a 0.5 m resolution) were utilized. However, Sentinel 2's resolution was quickly deemed insufficient for detecting the multitude of small landslides that had disrupted roads in numerous locations, and Planet's coverage was incomplete for the area in question. Consequently, the Emilia-Romagna Region opted for the acquisition of aerial imagery (0.2 m pixel, visible and NIR) over the entire affected area, which were completed and delivered as orthorectified images and WMS services at the end of June (five weeks after the event). Ever since, landslide and damage inventory were carried out manually based on these high-resolution aerial imageries.

The manual mapping process was both intensive and laborious. From June to July 2023, we focused our efforts on the regions most devastated by the event, particularly around roads and urban centers, in order to fulfill the Civil Protection Agency's urgent requirements for damage assessment during the emergency response. From August to November 2023, we extended our work to complete the landslide inventory for the entire affected area.

4.5 Methods

4.5.1 Study area

The assessment of automated mapping techniques took place in the Casola Valsenio municipality (Ravenna province), an area profoundly affected by the catastrophic events of May 2023. Casola Valsenio covers an area of 84 km² and is predominantly composed of the Marnoso-Arenacea Formation. This formation is stratigraphically divided into three members, differentiated by their varying ratios of sandstone to pelites, as depicted in Fig. 4.3. The structural setting of the area is consistent, characterized by a uniform homocline that gently dips toward the northeast. Under these conditions, the slopes within the Marnoso-Arenacea Formation exhibit a pronounced asymmetrical structure. This is clearly seen in the steep slope angles, which range from 30° to 70° where the rock layers angle towards the slope, compared to the gentler inclines of 10° to 20° found where the rock layers are parallel to the slope. Prior to the May 2023 event, the Geological Survey had documented 730 landslides within the area (Fig. 4.3). The majority of these were ancient, dormant rock-blocks slides on slopes parallel to the bedding planes.

The area is cloaked in thick forests, with extensive tree coverage. Farming is conducted on the valley bottom and the milder inclines. Casola Valsenio has a community of 2,650 residents and serves as a vital link connecting the Po Plain in the north to Tuscany in the south. In the course of the disaster, a significant landslide obstructed the main thoroughfare for weeks, causing considerable disruption for the locals and cutting off the southern part of the area. The landslides affected over 60 homes and damaged 180 roadways, inflicting harm that has yet to be completely remedied.

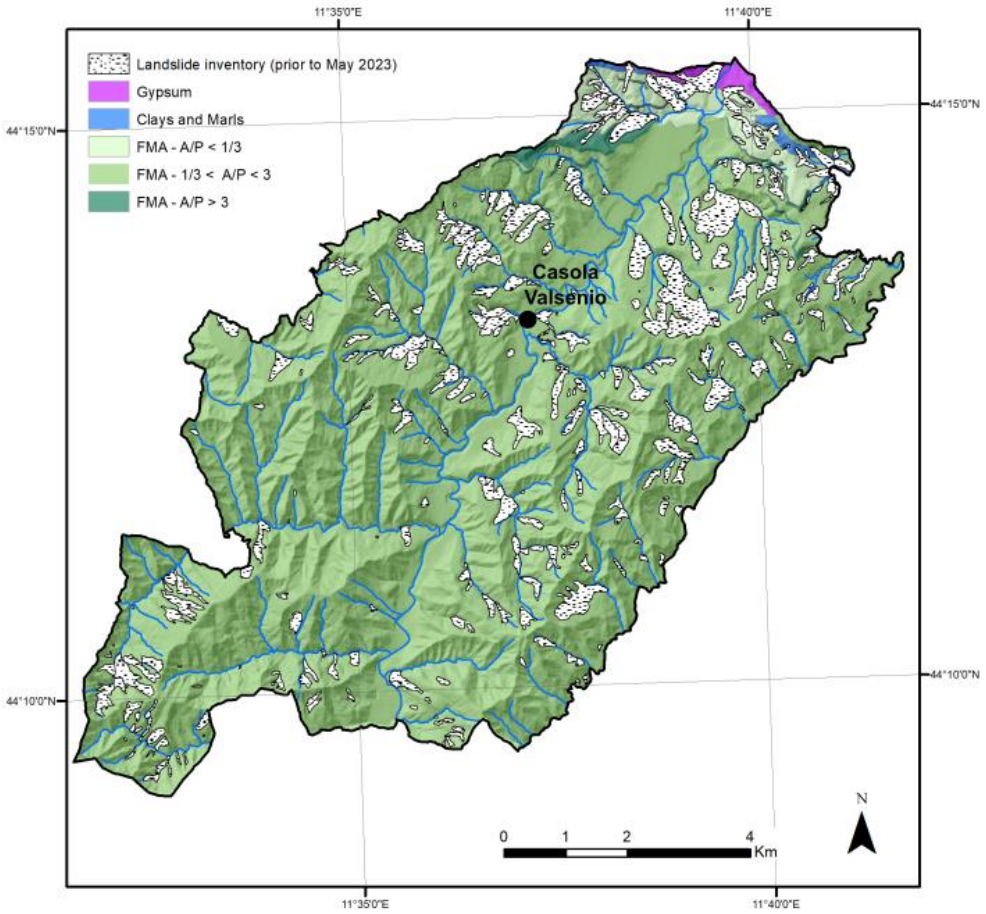


Figure 4.3. Schematic geological map of the municipality of Casola Valsenio (Province of Ravenna, Italy). FMA=Marnoso-Arenacea Formation, categorized into three distinct members based on the prevailing Areneties/Pelites (A/P) ratio.

4.5.2 Remote sensing images

In this study, we utilize three distinct types of remote sensing imagery to identify and map the landslides that occurred following the May 2023 rainfall events. Table 4.1 provides a summary of these images along with their specific characteristics.

Sentinel-2 images are provided by the European Space Agency and have a 10-meter resolution in the Blue, Green, Red, and near-infrared bands. The first post-event image with minimal cloud cover was captured on May 23, one week after the second rainfall event (dataset S3). Over the next month, eight more images with low cloud coverage were captured, forming an additional dataset (S4). Obtaining suitable pre-event images proved difficult due to persistent cloud cover throughout April 2023. Moreover, the last viable image from March 2023 did not accurately depict the vegetation condition during the event, as early Spring is typically marked by substantial vegetation growth and intense agricultural activities. To address this challenge, we

resorted to using two distinct pre-event datasets: S1, comprising multiple images from May and June 2022, and S2, consisting of one image taken exactly a year before the event.

AGEA, the Italian Agency for Agricultural Payments, captured high-resolution aerial photographs from 2009 to 2021, featuring a pixel size of 0.2 meters. These images are composed of 4 bands (RGB+NIR) and are instrumental in Italy for in-depth land use, crop health, and agricultural analysis. We utilized the 2020 AGEA image (dataset A1) to depict pre-event conditions at high resolution. This composite image, assembled from shots taken between April and July 2020, generally reflects the vegetation state during the event period, although there might be local variations owing to different agricultural practices.

CGR images (dataset C1) are the high-resolution aerial photographs taken after the event, commissioned by the Emilia-Romagna Region. These images feature 4 bands and have a pixel size of 0.2 meters. The imagery was collected using a Leica DMC III sensor mounted on a Cessna 402C aircraft, which flew at an altitude of 4718 meters above sea level. CGR photos were taken on May 23. The raw, non-orthorectified images were quickly made available within a week of acquisition. The orthorectification process was completed later, with the final images delivered on June 21.

Each dataset includes the near-infrared (NIR) spectral band, which proves particularly important for our analysis. Landslides often resulted in the removal of vegetation, and the NIR band is highly sensitive to such changes thanks to its ability to detect variations in chlorophyll and subsequent shifts in plant biomass. (Tucker 1979). Figure 4.4 highlights the importance of images resolutions (in both visible and NIR spectrums) for detecting landslides in our specific scenario. It is particularly noteworthy that the smallest landslides, those under approximately 400 m², can be identified in the high-resolution CGR images (Fig. 4.4a-b), whereas in the Sentinel-2 images, they are represented by a mere 4x4 pixels or less (Fig. 4.4c-d). Considering that most of the landslides were small or very small, high-resolution images were indispensable for their manual identification.

Table 4.1. Satellite and aerial imagery used for landslide mapping through manual and automated methods.

Dataset	Period	Image Source	Image Type	Spectral Bands	Pixel (m)	Date	Number of images
S1	Pre-event	Sentinel2	Satellite	RGB+NIR	10	5 May-30 June 2022	9
S2	Pre-event	Sentinel2	Satellite	RGB+NIR	10	13 May 2022	1
S3	Post-event	Sentinel2	Satellite	RGB+NIR	10	23 May 2023	1
S4	Post-event	Sentinel2	Satellite	RGB+NIR	10	23 May-30 July 2023	8
A1	Pre-event	AGEA	Aerial	RGB+NIR	0.2	April-July 2020	1
C1	Post-event	CGR	Aerial	RGB+NIR	0.2	23 May 2023	1

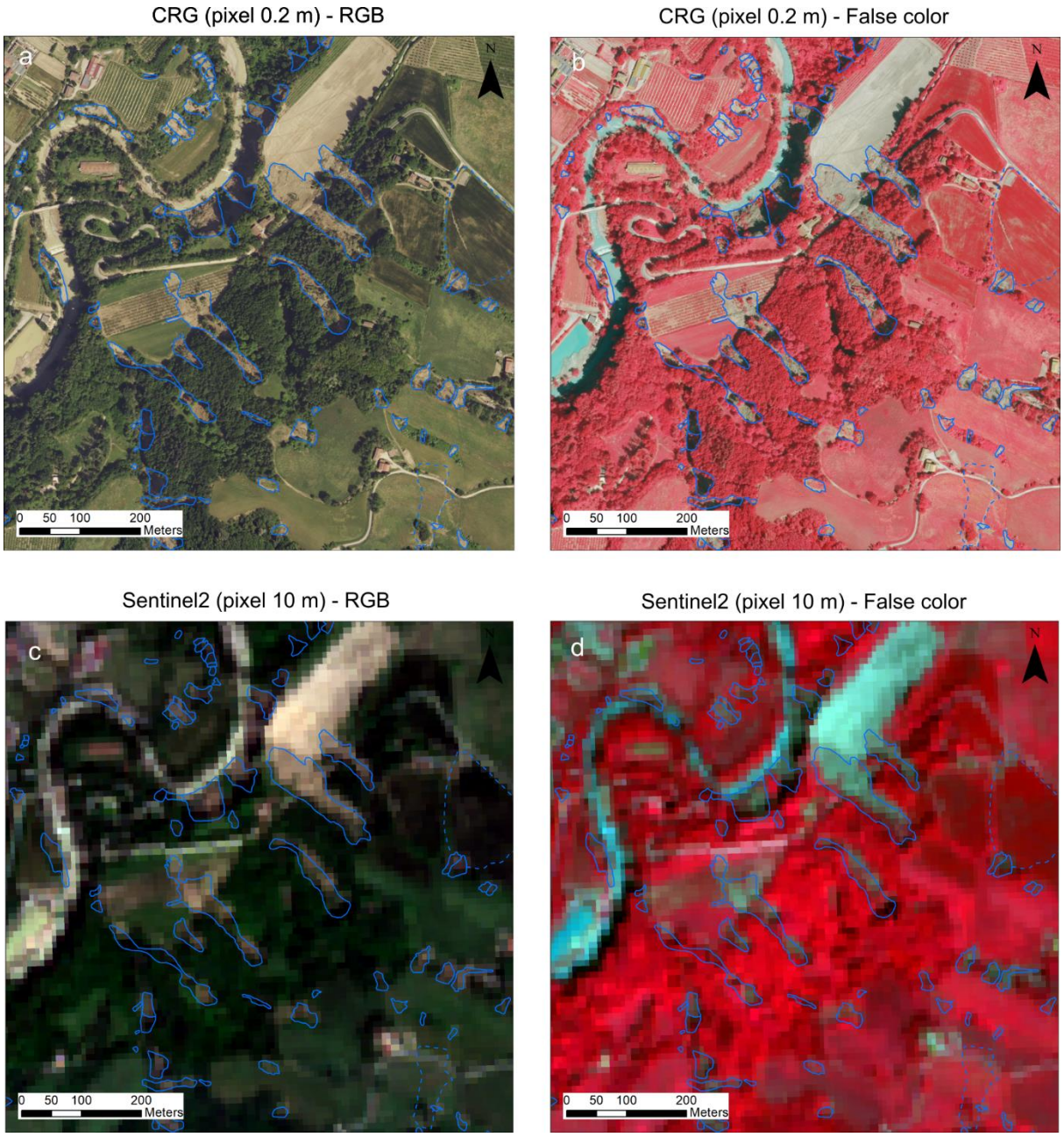


Figure 4.4. Examples of imagery from a small section of the study area illustrating various products. **A**) CGR aerial image with 0.2 m resolution in RGB; **b**) CGR aerial image with 0.2 m resolution in NIR false colors; **c**) Sentinel2 satellite image with 10 m resolution in RGB; **d**) Sentinel2 satellite image with 10 m resolution in NIR false colors.

4.5.3 Landslides manual mapping

A detailed inventory of landslides was manually compiled using high-resolution aerial imagery. We employed images from after the May 2023 event (0.2 m pixel dataset C1) and from before the event in 2020 (0.2 m pixel dataset A1). This task was performed by our team members who had conducted field inspections at the landslide sites soon after the event, particularly within

the first four weeks following May 16th. Their direct experience in the field was crucial for accurately identifying features of landslides in the aerial photographs.

The process involved visual inspection and digitization within a GIS framework, utilizing on-screen scales from 1:1000 to 1:200 for detailed examination. A 1x1 km reference grid was established to assist surveyors during inspections and to monitor the progress across the examined areas. Additionally, NDVI-change maps derived from Sentinel2 data served as a background layer to highlight potential land cover changes. Support for the surveyors also included NIR band imagery, hillshade renderings from a 5 m resolution DEM, and topographic maps to enrich the analysis and digitization process. Moreover, a 3D visualization tool was employed alongside the GIS setup to provide dynamic perspectives of the surveyed and mapped areas, enhancing the comprehensiveness of the landslide inventory.

Within the GIS framework, the outlines of landslides (as polygonal layers) and their presumed initiation points (as point layers) were digitized. The focus was exclusively on landslides that occurred during May 2023; those that were already present in the 2020 imagery and remained substantially unchanged by 2023 were not included in the inventory. For debris flows with several source areas, multiple initiation points were documented. The placement of these points, determined at the discretion of the operator, typically occurred in the highest parts of the detachment zones. The attribute table for the landslide layer featured a coding system to differentiate various landslide types, such as debris slide (DS), debris flow (DF), rock slide (RS), earth slide (ES), and earth flow (EF) (Cruden and Varnes, 1996). Additionally, an inventory of landslide-induced damages (as a point layer) was compiled by marking the midpoint of affected road segments or the centroid of damaged or at-risk buildings. Attributes were assigned to distinguish between roads and buildings, and for buildings, whether they were located directly within or proximate to a specified buffer zone of 30 m surrounding the landslide outline.

Most of the landslides occurred during May 2023 were first time failures, making them relatively straightforward to identify in aerial imagery. Debris flows (both channelized and non-channelized) and debris slides led to noticeable changes in land cover by clearing vegetation in the source areas and along the track and accumulating bare soil in deposition zones. Similarly, translational rock-block slides brought about considerable landscape changes, including clearly visible detachment scarps and extensive areas filled with large, block-shaped debris. Nonetheless, challenges in capturing all landslides in the inventory included the small scale of some events (just a few tens of square meters) and instances where landslides extended beneath canopy cover, obscuring them from aerial view. Additionally, for rock block-slides, aerial image reviews might have missed slope cracks indicative of early-stage landslide activity. Despite these limitations, we believe the manually compiled landslide inventory is nearly comprehensive. It's important to note that the accuracy of the manual mapping was significantly improved due to the largely homogeneous geological conditions. Almost the entire area is

dominated by a single geological formation, with consistent structural settings. This uniformity resulted in highly consistent and repetitive slope morphologies and landslide types. Field checks were also particularly effective in these relatively straightforward geological settings.

4.5.4 Automatic mapping techniques

The automatic mapping of landslides was conducted using two established techniques: the conventional NDVI (Normalized Difference Vegetation Index) method and the more advanced U-Net algorithm. Both methods are well-known and extensively referenced in the literature. Considering our goal of implementing automated landslide detection during emergency scenarios, we chose NDVI for its simplicity and ease of interpretation, and U-Net for its robust machine-learning capabilities, which have proven highly effective across various geological settings.

In the immediate aftermath of a disaster, there is no time to develop new methods or modify existing ones to achieve the highest possible statistical accuracy. Instead, the expectation is that established automated techniques will deliver reliable results quickly without the need for additional refinement. Therefore, both methods were applied in their standard forms without any significant modifications to the algorithms.

NDVI

The Normalized Difference Vegetation Index is a standard tool in remote sensing (Tucker 1979). Its effectiveness in detecting landslides stems from its capability to distinguish vegetated from non-vegetated surfaces by their reflectance characteristics. NDVI is computed using the equation (Kriegler et al. 1969):

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad [3.1]$$

Here, *NIR* refers to spectral reflectance in the near-infrared band, and *Red* represents visible red light. NDVI values fall between -1 and +1, with higher figures indicating denser vegetation. In landslide detection scenarios, a reduction in NDVI values is typically observed in affected areas, reflecting vegetation loss. This change is measured as ΔNDVI, calculated as the post-event NDVI minus the pre-event NDVI:

$$\Delta NDVI = NDVI_{post} - NDVI_{pre} \quad [3.2]$$

A positive $\Delta NDVI$ suggests an enhancement in vegetation greenness or health, whereas a negative value points to a decline. Establishing a universal NDVI change threshold for landslide detection is unfeasible due to the diversity of regional vegetation types, varying climatic

conditions, differences in landslide characteristics, and the varying resolutions of satellite imagery. Consequently, such thresholds are usually set through local calibration and validation, using known landslide events and corresponding NDVI changes (Yang et al. 2019; Qu et al. 2021).

The NDVI method was applied using different combinations of pre- and post-event images as outlined in Table 4.2. Cases NDVI-1 and NDVI-2 both employed Sentinel2 images but differed in their index calculation techniques. In NDVI-1, the NDVI rasters for both pre-event and post-event were generated by averaging NDVI values across numerous images over roughly two months to reduce noise and anomalies. In contrast, NDVI-2 calculated indices from just two distinct images, which improved the detection of vegetation changes caused by landslides. NDVI-3 used high-resolution aerial imagery with a resolution of 0.2 meters, providing much greater spatial detail that allows for the identification of smaller-scale vegetation changes not visible at 10-meter resolution. As described in section 4.5.6, the original 0.2-meter images were resampled to 1 meter to decrease the noise commonly found in high-resolution NDVI change maps.

The NDVI change method for landslide identification could be refined by integrating additional parameters such as slope, aspect, and others (e.g. Notti et al. 2023). Although these modifications have proved to be effective in specific contexts, we decided to use the conventional approach. This decision was driven by the practical challenges of implementing significant refinements in emergency situations. Enhancements in fact necessitate a detailed analysis of landslide-influencing factors, that are often not viable within the constrained timeframes typical of emergency scenarios.

Table 4.2. Summary of the cases considered in the automated landslide mapping. Dataset codes refer to Table 4.1.

Method	Case	Pre-event dataset	Post-event dataset	Pixel (m)	Description
NDVI	NDVI-1	S1	S4	10	Low resolution Sentinel2 images (10 m). Pre- and post-event NDVI maps created by averaging a series of images.
	NDVI-2	S2	S3	10	Low resolution Sentinel2 images (10 m). Pre- and post-event NDVI maps obtained by a single image.
	NDVI-3	A1	C1	1	High-resolution aerial images (0.2 m) resampled at 1 m.
U-Net	UNET-1	\	S3	10	Low resolution post-event Sentinel2 image (10 m).
	UNET-2	\	C1	2	High-resolution post- event aerial image (0.2 m) resampled at 2 m.

Table 4.3. Results of May 2023 landslides manual mapping in the municipality of Casola Valsenio.

Landslide type	Number [count]	Number [%]	Area > 10000 m^2 [count]	Total Area [km^2]	L/W ratio	Existing landslide area [%]
Rock Slides (RS)	111	2.0	17	0.629	0.99	30.9
Debris Slides (DS)	4224	76.7	0	1.733	1.07	7.1
Debris Flows (DF)	1076	19.5	17	1.809	2.10	20.2
Earth Flows (EF)	96	1.7	7	0.350	2.10	15.0
Total	**5507**	**100**	**41**	**4.521**		

U-Net

U-Net is a type of convolutional neural network (CNN) that was initially developed for biomedical image segmentation (Ronneberger et al. 2015). The network's architecture, resembling the letter "U," consists of a contracting (downsampling) path to capture the context of the input image and an expansive (upsampling) path for precise localization. This symmetry in the network, where layers in the downsampling path are mirrored in the upsampling path, helps in transferring information across different resolution scales. A crucial aspect of U-Net's design is its incorporation of "skip connections," which are direct links between layers of the same resolution in the downsampling and upsampling paths of the network. Skip connections enable the propagation of detailed local information, like textures or specific shapes, directly across the network. This information is crucial for precise segmentation, as it helps the network to accurately reconstruct the finer details of the image.

U-Net can be applied to landslide mapping by utilizing its segmentation capabilities to differentiate landslide areas from their surroundings in remote sensing images (Soares et al., 2020; Qi et al., 2020; Meena et al. 2022). The U-Net models are generally trained with images of various window dimensions (tiles) and sampled using different methods, including random and regular grid methods, with data augmentation techniques like random rotation and flips. Currently, in the application of U-Net for landslide mapping, there are no established general rules or standardized procedures regarding the number of tiles, their size, and other specific parameters. The optimal choices for these factors can vary significantly based on the specific characteristics of the study area, the resolution and type of the remote sensing data, and the specific objectives of the analysis. This lack of standardization reflects the ongoing development and adaptation of machine learning techniques like U-Net in the field of landslide mapping.

In our study, we utilized MATLAB to implement U-Net for landslide classification. The U-Net architecture was generated using the function `unetLayers([height, width, depth], numClasses)`. The parameters `height` and `width` define the size of the input

image, `depth` is the number of image channels, and `numClasses` specifies the number of classes for semantic segmentation. For the analysis we used both the Sentinel2 images (case UNET-1, Table 4.2) and the high-resolution CGR image (case UNET-2, Table 7). In both cases the `depth` is equal to 4, the number of spectral bands. We set `numClasses` to 2 for binary classification, distinguishing between landslide and non-landslide areas.

To configure a neural network like U-Net, several parameters known as hyperparameters must be defined. Hyperparameters are external configurations to the model and are not learned from the data. They play a critical role in controlling the behavior of the training process and the performance of the network. Key hyperparameters (Ronneberger et al. 2015) are detailed below.

- The `EncoderDepth` determines the number of encoder levels in the U-Net architecture. In our case, the network features a moderately deep encoder with 4 levels. This depth is a standard choice for various image segmentation tasks. Each encoder level is composed of two convolutional layers, followed by a ReLU activation function. This structure is consistent across all encoder levels, with max-pooling layers interspersed to diminish spatial dimensions and to deepen the feature channels.
- `NumFirstEncoderFilters` is set at 32, indicating that the initial layer of the network uses 32 filters. This number strikes a balance between the capacity to learn intricate features and the limitations of computational power and memory. Generally, 32 filters are adequate for simpler tasks or when standard computational resources are available.
- The `InitialLearnRate` sets initial size of steps in the optimization process. Setting this rate too high might prevent the network from converging or lead to a suboptimal solution, while too low a rate can result in excessively slow convergence. We experimented with different `InitialLearnRate` values, ranging from 0.0001 to 0.01.
- `SolverName` identifies the optimization algorithm used for updating the weights of the network. The Adam solver, known for Adaptive Moment Estimation, is a favored choice due to its efficiency and effectiveness in a broad spectrum of machine learning tasks.
- Lastly, `MaxEpochs` defines the number of complete passes through the training dataset by the network. The decision on the number of epochs was based on performance, concluding the training when the network ceased to show significant improvements in performance.

To enhance the model's capacity to generalize, we implemented artificial augmentation of the training images. This was achieved using MATLAB's `imageDataAugmenter` object, which was configured to perform the following transformations: random reflection in the horizontal and vertical direction; random rotation by an angle between -10 and +10 degrees; random horizontal and vertical translation (shift) by a value between -10 and +10 pixels; random horizontal and vertical shear with an angle between -2 and +2; random scaling by a factor

between 0.8 and 1.2. These augmentations greatly increased the diversity of the training dataset, making model stronger and more general.

4.5.5 Training and testing

Each method used for automated landslide mapping relies on a ground-truth dataset, which is essential for both training and testing the models. Typically, this dataset originates from a landslide map manually created by specialists. The dataset is then split into two distinct portions: a larger part is used for training the model, which involves tasks such as establishing the NDVI threshold or adjusting the U-Net hyperparameters, and a smaller part is reserved for testing purposes. Randomly splitting the dataset into training and testing sections, often in 80%-20% or 70%-30% ratios, is a common practice (Nava et al. 2022; Chandra et al. 2023). This strategy is effective under non-critical conditions, but it is unsuitable for emergency mapping scenarios. In emergencies, time is of the essence, and manually mapping landslides across 70%-80% of the area is not viable. An automated method should instead provide reliable results with minimal training conducted on just a fraction of the area. Therefore, in emergency contexts, the balance between training and testing datasets needs reevaluation.

To tackle this challenge, we adopted an unconventional approach. The study area was first segmented into 1x1 km grid tiles as shown in Fig. 3.5. Each 1 km² tile corresponds to an area of 100x100 pixels in the 10m Sentinel2 imagery, 500x500 pixels in the CGR imagery when resampled at 2m, and 1000x1000 pixels in the CGR imagery at 1m resolution. Generally, a single grid tile in these areas contains over 50 landslides, providing a reliable overview of landslide pattern. We then proceeded to train the NDVI and U-Net models using a limited number of tiles, specifically 1-2-4-6-8-10-12 tiles, aligning with the requirements for rapid mapping in emergency situations. Considering there are 86 tiles in Casola Valsenio, the extent of our training dataset ranged from a mere 1% (1 tile) to 14% (12 tiles) of the entire dataset. Testing was then carried out on the remaining tiles, leading to a varied training-to-testing ratio from 1%-99% to 16%-84%.

Ideally, training tiles should be distributed across the entire study area to encompass diverse conditions such as various terrain types, vegetation, and landslide characteristics. However, in practical scenarios, it is common for images to cover only a limited part of the area, often represented by several contiguous tiles. To address these varied scenarios, our methodology incorporated two different strategies for tile selection: a random layout and a clustered layout. In the random layout, training tiles are randomly selected across the study area. This approach aims to capture a wide range of conditions, thereby enhancing the statistical robustness and generalizability of the results. In contrast, the clustered layout involves grouping training tiles that are geographically close to each other. This approach is designed to simulate scenarios where post-event imagery is only available for a limited area and manual mapping are possible

only for specific local conditions. These selection strategies led to the creation of 14 unique training scenarios, evenly divided with 7 using a clustered layout and 7 using a random layout, as depicted in Figure 4.5.

The training process then varied for the two methods. In the NDVI method, training involves determining the ideal $\Delta NDVI$ value ($\Delta NDVI^{threshold}$) that effectively distinguishes between landslide and non-landslide pixels. To establish this threshold, we experimented with various $\Delta NDVI$ values within the theoretical range of -2 to +2. The effectiveness of each value in aligning the automatic classification with the ground truth in the training tiles was assessed using the F1-score. The F1-score is the harmonic mean of precision and recall (Tharwat 2018) and it is particularly suitable for models applied to unbalanced datasets like ours, where landslide pixels typically represent less than 10%-15% of the total. To evaluate the method's validity, the threshold determined from the training tiles was finally applied to the test tiles, and results assessed using the same statistical method. The F1-score calculated for the testing dataset serves as an indicator of the method's effectiveness.

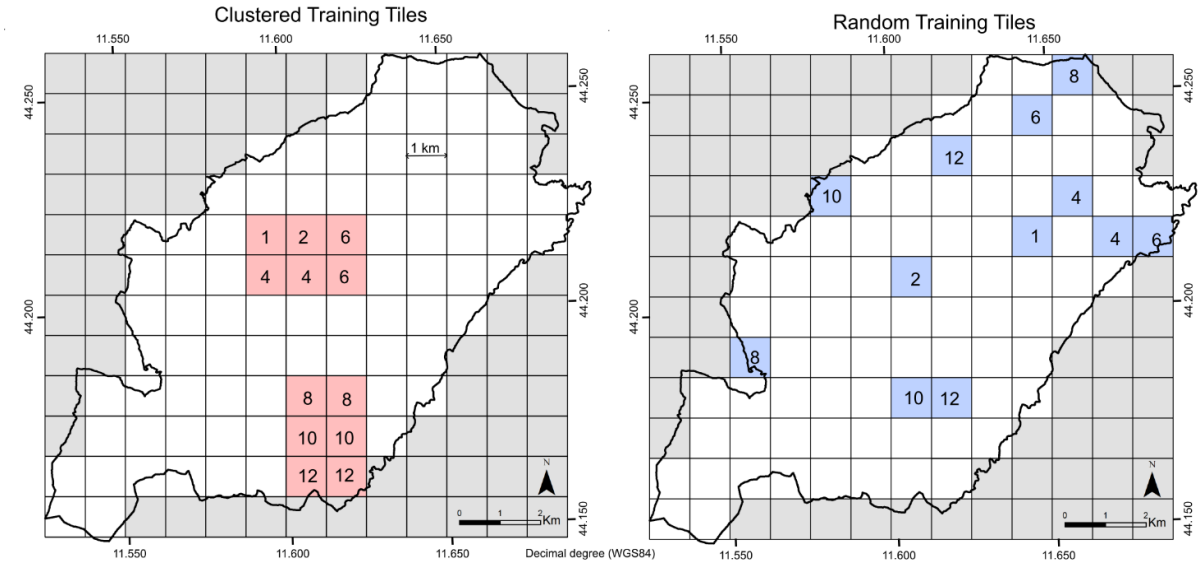


Figure 4.5. Location of the tiles utilized for training the automated mapping models in both clustered and randomly selected training scenarios. The numbers show the sequential inclusion of tiles in the analysis.

In the U-Net method, the training stage is fundamentally important. The neural network, during this phase, learns to recognize specific features, in our case the manually identified landslides, from the input imagery, which consists of a 4-band raster. This training is achieved through a process where batches of data are continuously fed into the network. The network makes predictions, and its weights are adjusted based on the calculated loss from these predictions. This cycle of prediction, adjustment, and feedback is repeated across multiple epochs, or

complete passes through the training dataset, until the network achieves satisfactory performance levels. Once trained, the U-Net model is then applied to test tiles to assess its effectiveness in real-world scenarios. Unlike the NDVI method, U-Net directly yields a binary classification output, categorizing each pixel as either a landslide (1) or a non-landslide (0). The accuracy and reliability of these predictions are evaluated again using the F1-score.

4.5.6 Sensitivity analysis

A detailed sensitivity analysis was performed on the previously described analysis process to evaluate the effects of input data on the results. The primary focus was on high-resolution images at 0.2 m (datasets A1 and C1 in Table 4.1). These images were crucial for manual mapping and allowed a precise detection and delineation of the landslides triggered by the event. However, incorporating such high-resolution images into automated mapping methods presents significant challenges.

In NDVI analyses, the use of high-resolution imagery can often introduce additional noise, primarily from shadows and topographical variations (Burgess et al. 1995; Yang et al. 2022). High-resolution sensors capture shadows cast by terrain features, vegetation, and built structures, which can lead to apparent fluctuations in NDVI that do not correspond to actual changes in vegetation Additionally, minor natural variations such as the movement of leaves and branches, or slight shifts in the sensor's angle, can further contribute to this noise. Thus, while finer resolution offers detailed spatial information that helps identify small-scale changes in vegetation, it does not necessarily enhance the overall quality of the outcomes. Conversely, U-Net generally performs better with high-resolution images. The higher spatial resolution enables the detection of critical details at finer scales, providing a richer set of features that enhance the model's ability to learn more intricate and detailed representations. However, using high-resolution images also presents challenges, including increased computational demands, greater memory requirements, and a heightened risk of overfitting, particularly when the available training dataset is limited.

To tackle these challenges, we carried out a sequence of analyses by modifying the pixel size of the high-resolution images. We resampled the original 0.2 m images to resolutions of 0.5, 1, 2, 5, and 10 meters using bilinear interpolation. This resampling was performed on both pre-event and post-event images to generate lower-resolution NDVI and RGB rasters. The resampling process was applied directly to the original spectral bands to maintain the integrity of the spectral data. We then conducted NDVI and U-Net analyses using the same tiles for both training and testing, but with varying image resolutions. For each combination, we compared the F1-score, computational time, and memory requirements. The raster resolutions specified in Table 4.2 (1 m for NDVI-3 and 2 m for UNET-2) were identified as the optimal balance and selected as reference cases for this study.

Specifically, for the NDVI method, a resolution of 1 m strikes an effective balance between accurately defining landslide boundaries and mitigating the salt-and-pepper effect observed in ultra-high resolution NDVI change maps. The full resolution of 0.2 m provides in fact precise delineation of areas where vegetation has been cleared, but it also leads to the creation of numerous small polygons in forested areas. These polygons, generated due to shadow noise, are mistakenly identified as landslides.

The 2 m resolution adopted for U-Net analysis mostly derives from the challenges in memory allocation with high-resolution data. In our case, a 1 km² training tile corresponds to an array size of 500x500 pixels for a downsampled 2 m image, and 5000x5000 pixels for a full-resolution 0.2 m image, requiring memory sizes of 4 MB and 400 MB respectively (assuming 4 bands and a depth of 32 bits). When using deep learning models like U-Net, the memory consumption during training significantly exceeds the space needed merely to store the input images, due to additional requirements such as storing augmented image versions and managing the weights and gradients of various layers. As a result, a typical personal computer equipped with 12GB of VRAM struggles to handle even a single full-resolution tile, while it can easily train a model with 12 tiles at 2 m resolution. Moreover, the computational time increases in a non-linear manner as the input image size grows, which becomes a significant issue when timely results are needed. For instance, using a GeForce RTX 3030 GPU, training a 1 km² tile for 100 epochs takes about 1 minute at 2 m resolution, but it extends to 5 days at 0.2 m resolution. Given these constraints, running a U-Net model at full resolution has proven to be impractical with the time and resources available.

4.6 Results

4.6.1 Landslides manual mapping

On a regional scale, manual mapping has cataloged over 80,000 landslides, encompassing an area greater than 1,000 square kilometers. Predominantly, these landslides consist of debris slides, which make up nearly 50% of the total, and debris flows, which account for over 20%. Rock-block slides, while less common (constituting less than 1%), are notably significant due to their substantial size and depth, which often result in severe damage when they intersect with roads and buildings. In the Romagna region, which stretches from Imola to Cesena and features the prevalent Marnoso-Arenacea Formation, landslides recorded in May 2023 have been documented with an extraordinary density exceeding 200 landslides per square kilometer (see Fig. 4.1a). A comprehensive description of this regional landslide inventory has recently been submitted for publication (Berti et al., 2024 " RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event ").

The municipality of Casola Valsenio (Senio valley, Imola, Province of Bologna) records one of the highest landslide densities (Fig. 4.6). This area alone is responsible for over 100 rock-block

slides (RS), constituting a significant portion of the total RS found throughout the entire Romagna region, which shares similar geological settings. Debris slides (DS) and debris flows (DF) constitute nearly 95% of the mapped events (Tab. 4.4), while the remainder consists of earth flows (EF). These latter types are primarily found in the northern sector of the area, where Pliocene clay formations are exposed.

The elongation ratio distinctly varies among the types of landslides mapped, from approximately 1 for rock and debris slides (RS, DS) to over 2 for debris and earth flows (Tab. 4.4). Only about 20% of the mapped landslides were previously recognized in the official Italian landslide inventory (i.e., the IFFI database). Moreover, in those instances, the majority should not be seen as reactivations of pre-existing landslides but rather as initial failures that propagated atop former landslide deposits. In Casola Valsenio, landslides notably disrupted roads and homes, marking it as a critical area during the May 2023 crisis: a total of 249 damage points related to landslides were recorded (Fig. 4.6).

Table 4.4. Effectiveness of automated mapping models across scenarios outlined in Table 4.2. F1-score values represent the highest score achieved on the validation dataset. 'Threshold' denotes the optimal NDVI Change threshold identified.

Case	Training tiles distribution	Parameter	Number of training tiles						
			1	2	4	6	8	10	12
NDVI-1	Clustered	F1-score	0.49	0.51	0.51	0.52	0.51	0.52	**0.52**
		Threshold	*-0.1*	*-0.12*	*-0.12*	*-0.12*	*-0.12*	*-0.14*	*-0.14*
NDVI-2		F1-score	0.4	0.42	0.45	0.45	0.46	0.45	**0.45**
		Threshold	*-0.08*	*-0.1*	*-0.14*	*-0.16*	*-0.16*	*-0.16*	*-0.16*
NDVI-3		F1-score	0.46	0.45	0.46	0.46	0.45	0.45	**0.45**
		Threshold	*-0.34*	*-0.32*	*-0.35*	*-0.34*	*-0.34*	*-0.35*	*-0.35*
NDVI-1	Random	F1-score	0.49	0.51	0.51	0.52	0.51	0.52	0.52
		Threshold	*-0.1*	*-0.12*	*-0.16*	*-0.16*	*-0.16*	*-0.14*	*-0.14*
NDVI-2		F1-score	0.43	0.42	0.45	0.45	0.45	0.45	0.45
		Threshold	*-0.2*	*-0.22*	*-0.18*	*-0.18*	*-0.17*	*-0.17*	*-0.17*
NDVI-3		F1-score	0.36	0.41	0.45	0.45	0.44	0.46	0.46
		Threshold	*-0.47*	*-0.43*	*-0.37*	*-0.37*	*-0.41*	*-0.37*	*-0.37*
UNET-1	Clustered	F1-score	0.33	0.35	0.35	0.32	0.36	0.36	**0.35**
UNET-2		F1-score	0.47	0.50	0.56	0.60	0.59	0.64	**0.64**
UNET-1	Random	F1-score	0.20	0.26	0.36	0.32	0.37	0.37	0.37
UNET-2		F1-score	0.46	0.51	0.58	0.60	0.58	0.62	**0.58**

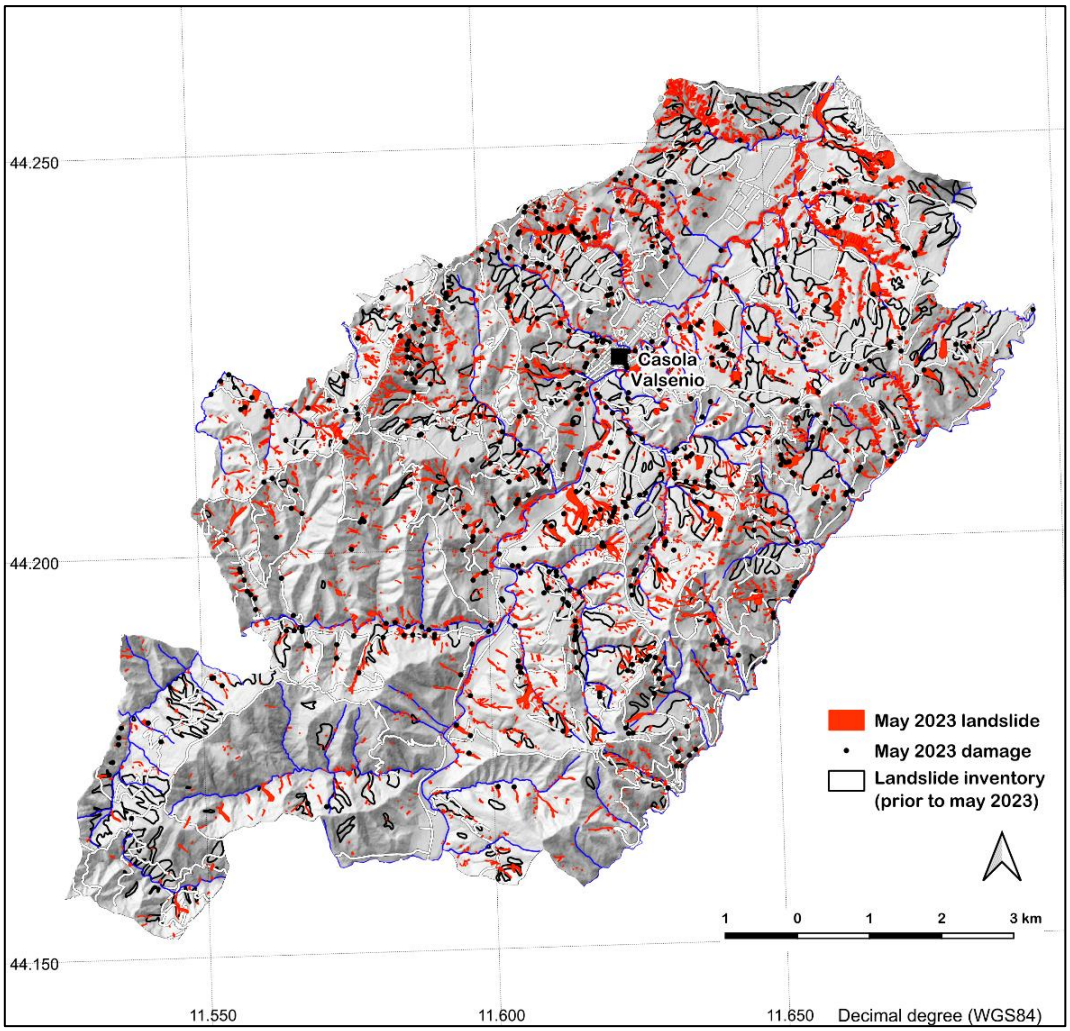


Figure 4.6. Map displaying landslides identified through manual analysis of high-resolution aerial imagery (0.2 m resolution) from the May 2023 event.

4.6.2 Landslides automatic mapping

Figure 4.7 summarizes the outcomes of the two different automatic mapping techniques applied to our study area. The figure displays charts with the number of training tiles in the x-axis and the F1-scores of the testing tiles in the y-axis. These scores provide a statistical measure of how well the automated mapping compares to the manual mapping. This is done for both scenarios where training tiles are either clustered or randomly distributed. Detailed numerical data and the optimal ΔNDVI thresholds can be found in Table 4.4.

From these results, we can observe the following points:

- The U-Net method outperforms the NDVI approach when using high-resolution images (UNET-2, F1=0.64), with F1-scores up to 30% higher than those achieved by the best NDVI map (NDVI-1, F1=0.52).

- U-Net's effectiveness significantly drops with lower-resolution Sentinel2 images (UNET-1, F1=0.35), bringing its performance below that of the NDVI approach.
- NDVI's performance is more influenced by the choice of pre- and post-event images rather than image resolution. The best NDVI results are obtained using Sentinel2 images where pre- and post-event scenes are averaged over a span of about two months (NDVI-1). In contrast, employing a single pre- and post-event image (NDVI-2 and NDVI-3) leads to considerably lower NDVI scores (F1=0.45 in both cases). Unexpectedly, reducing the pixel size from 10m to 1m does not lead to an improvement in the F1-scores.
- The optimal NDVI threshold varies with image resolution, showing values around -0.14 to -0.12 for Sentinel 2 images (10 m resolution) and between -0.40 to -0.34 for high-resolution images downsampled to 1 m (Table 4.4).
- The arrangement of training tiles, whether clustered or randomly placed, seems to have a negligible impact on the accuracy. Notable differences in performance are only seen when U-Net is trained with a very limited number of tiles (less than 4).
- A relatively small number of tiles, covering about 2-3% of the total study area, is sufficient to achieve the highest score for each method.

A closer examination of the landslide maps produced by the two distinct techniques highlights differences that the F1-scores do not fully capture. Figure 3.8 illustrates these differences in a specific part of the Casola Valsenio study area, selected for its diverse landscape features such as farmlands, wild terrains, wooded slopes, and landslides varying in shapes and sizes. The prediction maps were generated using 12 training tiles (see circled symbols in Fig. 4.7).

The NDVI method (NDVI-1/2/3) successfully detects most landslides, but it also leads to some false positives and false negatives. In particular, agricultural areas often result in false positives, especially when post-event imagery shows plowed fields or significantly less green vegetation compared to pre-event images (as shown for example in Fig. 4.8e under FP). Reducing these false positives remains challenging, even when attempting to use slope angle to exclude flatter regions, because many landslides started on steep slopes but propagated over flatter areas. False negatives occur less frequently, typically in situations where a landslide occurs on an already bare slope (FN in Fig. 4.8e), or when landslide debris spreads through a forest without markedly altering the vegetation.

The spatial fragmentation of the NDVI results, evident in Fig. 4.8cdef and Fig. 4.9c, poses additional challenges for landslide classification and risk assessment. Fragmentation often results in isolated clusters of pixels being identified as potential landslides, instead of continuous and coherent landslide bodies. This can complicate the interpretation of results, as small, scattered clusters may represent minor debris or noise rather than meaningful landslide

features. For operational purposes, such fragmentation could lead to misestimation of affected areas, impacting subsequent risk analysis and emergency planning.

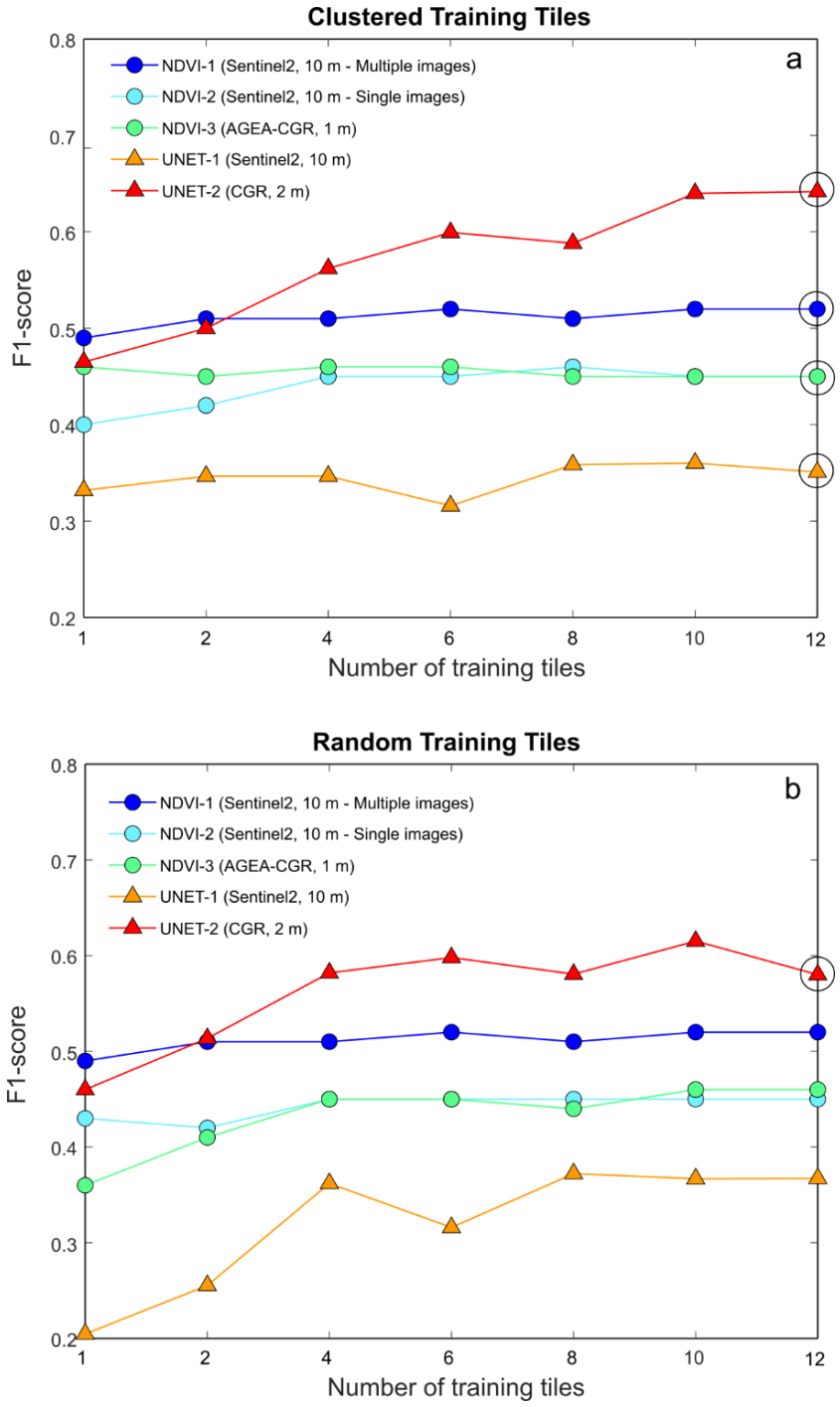


Figure 4.7. F1-score values for different automated mapping methods as a function of the number of training tiles. a) Clustered distribution of the training tiles; b) Random distribution of the training tiles. The circled symbols mark the cases depicted in Fig. 4.8.

The U-Net method applied to high-resolution images (UNET-2) exhibits high accuracy with a low rate of misclassified pixels (Fig. 4.8g-h). However, understanding why these misclassifications occur is difficult. Unlike simpler methods like NDVI, the inner mechanisms of U-Net are not easily interpretable, and to comprehend why U-Net incorrectly classifies certain pixels, one should investigate the neural network's layers. This involves understanding how it processes and identifies features, which is not practically feasible. In our case, U-Net primarily exhibited a tendency towards false positives rather than false negatives. These false positives typically include areas experiencing soil erosion (FP in Fig. 4.8g), which are slopes impacted by rainfall but not landslides, as well as sections of roads covered partially by mud or dust, or moist areas in plowed agricultural fields that resemble landslides in shape. Another notable source of false positives are backyards and the roofs of houses. Indeed, some false positives are areas where manual classification was ambiguous, indicating that they might actually represent real landslides.

Although U-Net results (e.g., Fig. 4.8g-h) exhibit less spatial fragmentation than NDVI, some degree of fragmentation is still present. Fragmented detection is more pronounced in areas with overlapping vegetation or complex topography, where the detachment zone and debris accumulation may be classified as distinct events. This is especially evident in the case of rock-block slides (Fig. 4.9d), where the spatial separation of detachment and debris areas further complicates the classification process. Such fragmentation can undermine the ability to provide a cohesive representation of the landslide body, potentially leading to underestimation of its scale or misinterpretation of its impact during risk assessments.

The bad results of U-Net when used with Sentinel2 images (UNET-1, Fig. 4.8f) is largely due to the limited resolution of these images, which hinders the detection of smaller landslides. Landslides that are less than approximately 400-500 m² in size, roughly equivalent to a 2x2 pixel area, are challenging to discern even with the naked eye. The training method of U-Net, which relies on pattern recognition and contextual understanding similar to human cognitive processes, also encounters limitations under these conditions. Of course, this is not a universal limitation. U-Net has shown its effectiveness in working with Sentinel2 images when dealing with larger landslides that occupy a more substantial number of pixels (e.g. Chen et al. 2023).

Regarding the ability to identify the various landslide types, both methods effectively detect debris slides (DS) and debris flows (DF), but struggle with mapping rock-block slides (RS). As expected, the first two types are easier to spot as they usually stripped away vegetation, in contrast to rock-block slides, which often leaved a substantial part of the sliding block with its vegetation intact. In such cases (see an example in Fig. 4.9), both NDVI and U-Net methods detect the detachment area and the fragmented debris at the toe as distinct landslide events, yet they fall short in identifying the full extent of the landslide body.

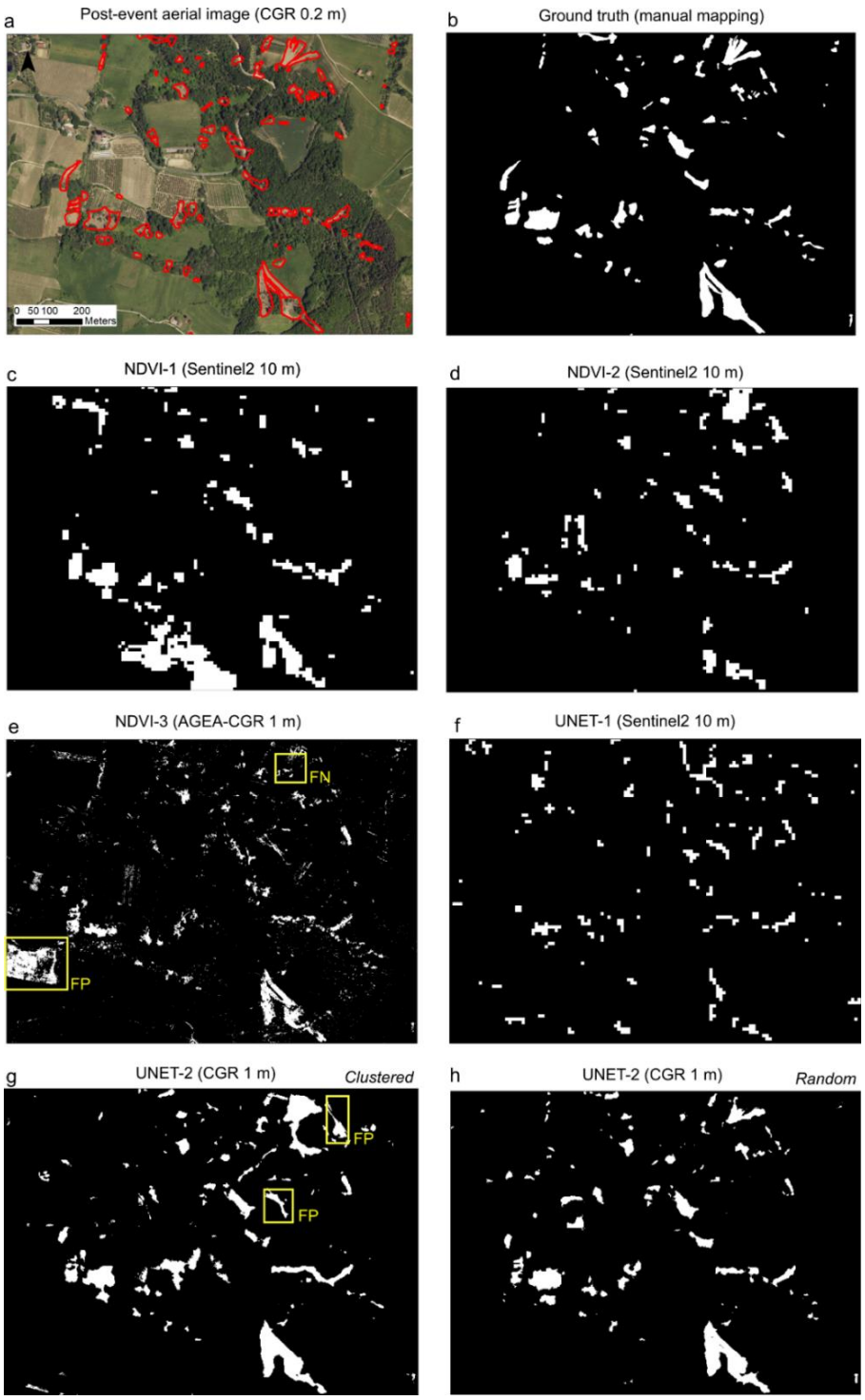


Figure 4.8. Evaluating the performance of various automated methods in a selected area of Casola Valsenio. **a-b**) Baseline data established through manual mapping; **c**) Landslide map generated using the NDVI change method with an average from multiple Sentinel2 images; **d**) Similar to (c) but with NDVI change calculated using a single image; **e**) Landslide map derived from NDVI change in high-resolution AGEA-CGR images; **f**) Landslide map produced using UNET on Sentinel2 images; **g-h**) Landslide maps created with UNET applied to CGR images, showing outcomes for both clustered (g) and randomly distributed (h) training tiles. FP=False Positive; FN=False Negative (see text).

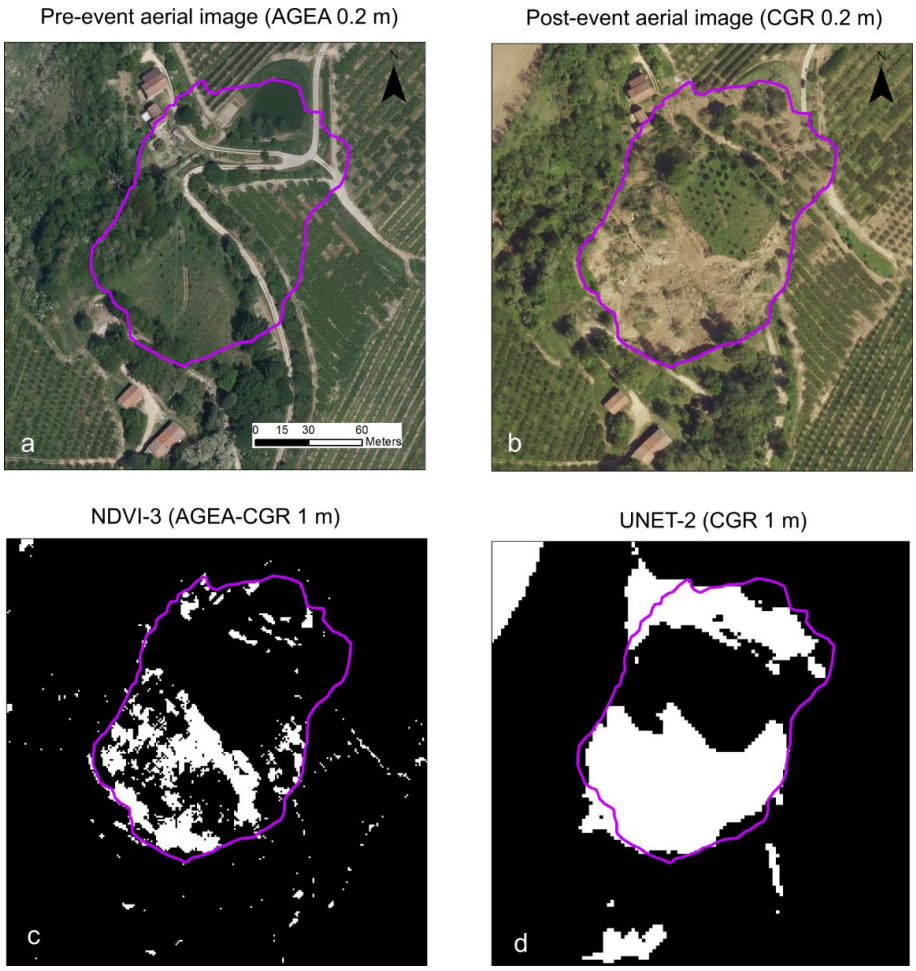


Figure 4.9. Challenges in automatically mapping rock-block slides. **a-b**) High-resolution images before and after a typical rock-block slide in the study area; **c-d**) Automated landslide maps generated from high-resolution imagery using the NDVI change method (c) and the UNET algorithm (d).

Figure 4.10 provides an additional comparison between the two methodologies. The figure represents the 1-kilometer tiles using a color scheme based on the F1-scores, computed for high-resolution images in the case of 12 clustered training tiles. The colors indicate four accuracy levels in the automated landslide maps, spanning from very high (F1>0.75) to very low (F1<0.25). Tiles utilized for training are distinctly marked with a thick black border. The data shows that the U-Net model, when trained using high-resolution images (UNET-2), achieves F1-scores above 0.5 in 75 of 86 tiles (approximately 87%). In contrast, the NDVI-3 model achieves similar scores in only 32 of 86 tiles (about 37%). In the U-Net model, the tiles with the lowest F1-scores present a simultaneous combination of low True Positives and high False Positives. This pattern is often observed in tiles with a sparse presence of landslides, where the few positive pixels do not align well with the actual occurrences. Conversely, in the NDVI model, the lowest F1-scores are primarily due to a high rate of False Positives. These inaccuracies tend to be distributed randomly across the tiles and are often associated with

agricultural activities. As expected, the training tiles (marked with a thick black border in Fig. 4.10) exhibit higher F1 values, particularly for U-Net, which is designed to identify complex data patterns. This effect is more subdued for NDVI, as this method mainly tracks changes in vegetation and is less sensitive to localized conditions.

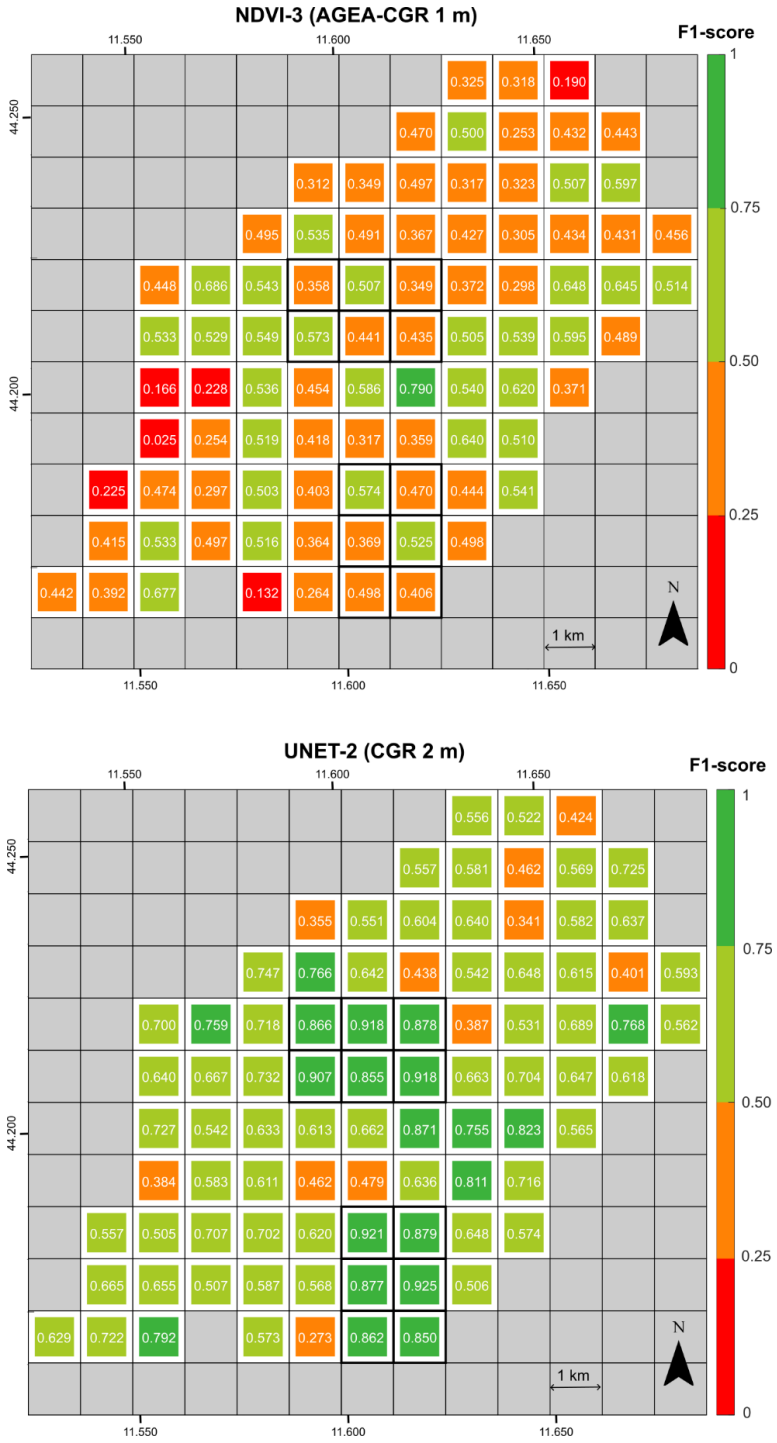


Figure 4.10. Performance comparison of the landslide map produced by NDVI change (upper) versus UNET (lower) across Casola Valsenio. Each 1km² tile is color-coded based on the F1-score, as indicated by the colorbar.

4.7 Discussion

4.7.1 Which method is best for automated mapping?

When strictly assessed on a statistical basis, U-Net clearly outperforms NDVI. In our study area, U-Net consistently achieves higher F1-scores, with a clear distinction in accuracy when compared to actual data. A notable strength of U-Net is its ability to differentiate between landslides and plowed fields, something that NDVI often confuses. Additionally, U-Net is more effective at spotting landslides on bare slopes, which NDVI frequently overlooks. This enhanced accuracy is attributed to the capabilities of convolutional neural networks, which, unlike the NDVI method that depends solely on pixel characteristics, analyze both the shape and internal patterns of objects. Furthermore, U-Net is proficient at handling high-resolution images, resulting in high-quality prediction raster.

The two methods, however, needs different skill levels, a crucial factor in emergency situations. NDVI is straightforward and simple, requiring no specialized hardware, with results that are easy to interpret and refine. For example, false positives in agricultural fields could be easily recognizable by their shape, and noisy pixel clusters could be filtered out to enhance map quality. U-Net, conversely, demands suitable computational resources and careful tuning of the model's hyperparameters to achieve optimal outcomes. This tuning process is complex and calls for an in-depth knowledge of the technique. Many technicians in civil protection agencies may lack the training for this task. Moreover, U-Net's outcomes tend to be more opaque or "black box" in nature, posing challenges in refining them without a thorough understanding of the method.

In the Romagna 2023 event, U-Net emerged as the best choice primarily because of the poor performance of NDVI with high-resolution images. This was largely due to misclassifications arising from agricultural activities, along with noise from shadows and vegetation changes. Generally, NDVI performs better when assessed over a time period rather than from single-event images (Martinez et al. 2021; Pu et al. 2022). Also our study revealed that using a series of Sentinel2 images (NDVI-1, F1=0.52) notably enhanced the NDVI's F1-score compared to single-image usage (NDVI-2, F1=0.45). However, this approach is often impractical with high-resolution images, due to their limited availability before and after disasters. Consequently, NDVI might face difficulties in emergency situations, particularly when dealing with smaller landslides and high-resolutions images are necessary.

4.7.2 Can automated mapping replace manual work?

In emergency scenarios, landslide mapping primarily serves two key purposes: i) provide a first look at the disaster's magnitude, which includes counting landslides, detailing their spatial layout, and identifying the affected regions; ii) detect the vulnerable elements impacted by the

landslides and provide a first estimate of the economic costs involved. The first goal is crucial for informing media, influencing political decisions, and raising public awareness about the event. However, the second aim is of even greater importance to the communities directly affected by the disaster, as it is crucial for accelerating the recovery process. In both cases, the precision of the landslide map is of paramount importance.

After the Romagna 2023 disaster, it was clear that manual mapping of landslides was necessary because automated NDVI maps were not detailed enough for important tasks like spotting infrastructure affected by landslides. As conditions improved two months later, we started exploring the application of U-Net technology. We are now in a position to assess whether advanced methods like U-Net can effectively enhance mapping efforts in emergency scenarios. This assessment involves (1) comparing the overlap of landslides identified manually to those identified by automated methods, and (2) measuring the accuracy of automated methods in identifying damaged buildings and roads by comparing these findings to manual inspections.

Regarding the first evaluation, our findings indicate that U-Net trained on high-resolution images (UNET-2), and NDVI trained on Sentinel2 images (NDVI-1), both deliver satisfactory results. Notably, U-Net achieves a good F1-score of 0.64 with minimal training involving only 12 tiles of 1 km² each. Automatic mapping methods, however, fall short in identifying damaged structures, as shown in Figure 4.11. This figure compares the counts of houses and roads affected by landslides, as determined through manual mapping, against the numbers detected automatically by overlaying NDVI and U-Net maps with layers indicating vulnerable elements. In Casola Valsenio, manual mapping identified 64 houses and 185 roads impacted by landslides. U-Net, utilizing high-resolution imagery and trained on 12 tiles, was able to detect a considerable portion of the damage to roads and buildings (65% and 72%, respectively). However, it also generated a significant number of false positives, particularly with buildings, often misclassifying roof tops as landslide areas. On the other hand, the NDVI mapping using Sentinel-2 images (NDVI-1) had fewer issues with false positives, but its accuracy in correctly identifying damaged buildings dropped to 49%. For both methods, precision (measured as the ratio of True Positives to the total Positives identified) is less than 0.52 for both roads and buildings. Additionally, while both methods tend to overestimate damages, they also generate several False Negatives. This suggests that their predictive accuracy is somewhat limited for this particular task.

Based on these findings, we must conclude that in our case automated methods cannot fully replace manual mapping. The precision in detecting damages is in fact too low and does not meet our specific requirements. Nonetheless, there are two considerations that warrant further discussion on this conclusion.

The first point to consider is the potential for improving our automated mapping. Enhancements to the NDVI and U-Net methodologies could be achieved through incorporating additional data

layers, applying post-processing filters, or tweaking the underlying algorithms. Many studies have documented the effectiveness of these approaches (Soares et al. 2020; Su et al. 2021; Bhuyan et al. 2023). As an example, consider how the results change when additional layers are incorporated into the U-Net model. Figure 4.11 presents two scenarios, UNET-A and UNET-B, illustrating the enhanced damage detection capabilities of a modified U-Net version that includes a fifth band of ΔNDVI data (sourced from CGR-AGEA and Sentinel2, respectively). Incorporating the ΔNDVI data should assist the network in differentiating between actual landslides, bare soils, and areas of similar colors, such as rooftops, thereby decreasing the number of false positives. Indeed, with this addition, the networks achieve a remarkable F1-scores of 0.68, and there is a reduction in false positives. The precision also improves compared to the 4-bands U-Net, attaining 0.46-0.55 for buildings and 0.65-0.69 for roads**.** While it is true that NDVI is calculated using R and NIR, which are already part of the input data, its inclusion as a fifth band provides a non-linear synthetic representation that can improve the model's ability to identify patterns in complex contexts. This observation is supported by the results shown in Chapter 5 (Table 5.2), where the inclusion of NDVI improves F1 and IoU scores for the Casola Valsenio municipality. These findings suggest that NDVI contributes complementary information, which positively impacts prediction performance, despite the potential overlap with R and NIR. Further optimization techniques were not pursued in our study because they are not practical in emergency contexts, where rapid response is crucial and there is often a lack of detailed ground-truth data. Nonetheless, it is conceivable that more advanced automation techniques could yield results closer to the ideal. This also applies to the use of full-resolution images. We had to downgrade the original images from 0.2 m to 2 m resolution to conduct the analysis within a reasonable timeframe and ensure a consistent sensitivity analysis. However, it is possible that with high-speed parallel computing, full-resolution images could be processed to achieve better outcomes.

The second point touches on a more "philosophical" question regarding the notion of ground truth. In our study, we treated the outcome of manual mapping as an accurate depiction of the actual landslides caused by the events, following common practice. However, it is important to acknowledge that manual mapping comes with its own level of subjectivity, casting some doubt on what we accept as ground truth. Research has indicated that variability among different mappers can introduce a significant subjectivity when delineating landslides from high-resolution aerial photographs (Guzzetti et al. 2012). This inherent limitation also affects the accuracy of what we consider ground truth. In our efforts to map landslides, we frequently marked areas where debris had dispersed among trees to capture the full extent of the landslide. This process required interpretation and led to the inclusion of areas with vegetation within the landslide's boundaries. Moreover, differentiating between shallow landslides and soil erosion, or identifying landslides that moved debris over hard rock cliffs, presented difficulties in certain cases.

It is unrealistic to assume that automated techniques can replicate such expert judgments. While methods like U-Net are designed to identify patterns from provided data, they naturally fall short of mimicking the intricate insights of a seasoned human mapper. Refining these methods could serve to narrow the divide between human expertise and machine precision in mapping, yet it is essential to acknowledge that manual mapping is susceptible to mistakes and inaccuracies. Striking a careful equilibrium between these elements is difficult, especially in complex mapping situations.

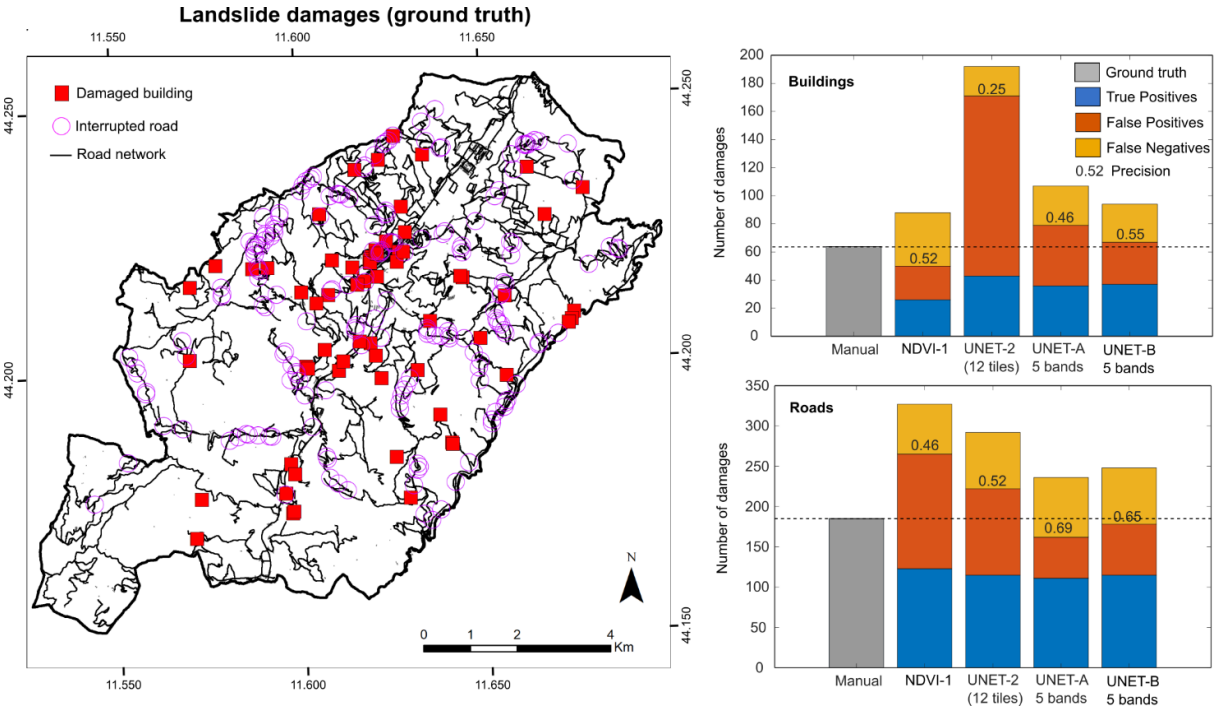


Figure 4.11. Landslide-induced damages within the Casola Valsenio area. The right-side map illustrates the positions of damaged buildings and roads identified through post-event aerial image analysis. On the left, histograms show the comparison between the actual number of damaged buildings (upper) and roads (lower) versus predictions made by various automated techniques. NDVI-1 represents the NDVI change from Sentinel2 images; UNET-2 denotes UNET application to CGR images; UNET-A/B are variations of UNET-2 incorporating an additional fifth band from NDVI changes using AGEA-CGR and Sentinel2 images respectively.

4.7.3 Lessons learned

Even the most skilled and experienced civil protection agencies, staffed with highly trained professionals, can encounter significant challenges during extraordinary rainfall events. This was exactly the case in our study. The Civil Protection Agency of the Emilia-Romagna Region, despite its extensive expertise and state of readiness, faced considerable difficulties in managing the scale and immediacy of the crisis. One of the most critical challenges immediately following

the disaster was the identification of impacted areas. Although there was support from the Copernicus Emergency Management Service and the National Civil Protection Agency (as mention in section 4.4.3), this crucial task was largely performed through manual efforts, requiring a great deal of work.

The extensive testing carried out in this study points to a feasible strategy for improving crisis management in future events. A suggested protocol for operational actions could be:

1) Quickly collect remote sensing imagery with high resolution immediately following the event. These images should ideally have pixel sizes ranging from 1-3 meters and encompass 4 bands (RGB+IR) to evaluate vegetation conditions more accurately.
2) Perform field surveys in selected areas to understand the specific landslide types and the general instability processes. These direct observations are crucial for accurate landslide mapping from the remote sensing data.
3) Manually map landslides in small, representative sections of the area (training tiles) that encompass the variety of landslide types and their geographical spread. Our findings indicate that mapping 2-3% of the total area could be enough for effectively training an automatic model during an emergency.
4) Use an automatic method to map landslides across the remaining area. U-Net, in our experience, was more accurate than traditional NDVI methods, though it requires a more complex training and validation process.
5) Overlay the landslide map with the vulnerable elements (roads, buildings, and infrastructures) to obtain a preliminary location of damages
6) Carefully check automated mapping results. Manual verification of samples is critical to identify errors and refine the automatic techniques

Key to this process is accessing high-resolution images after the event. While aerial photographs may not always be obtainable, satellite imagery of adequate resolution can be sourced or scheduled from platforms such as WorldView-3 and WorldView-4, GeoEye-1, and Pleiades-1A and Pleiades-1B. During the Romagna 2023 event, the mapping products rapidly generated from the Copernicus Emergency Service had limited utility since they were not tailored for the specific situation. Access to the original high-resolution satellite images (Pleiades-1A/B, with a 0.5 m resolution) that were used for analysis would have significantly aided in developing a dedicated mapping approach. We hope such images will be more readily available in future scenarios.

A final point to consider is communication. Managing the flow of information efficiently is vital at all stages of the process, encompassing data gathering, organization, and dissemination

from the community to local authorities and up to the central civil protection system. In our case, a major challenge in data collection was due to local authorities' limited experience in handling and disseminating data in GIS format and their difficulties with managing spatial data on field devices or smartphones. This gap in digital proficiency hindered not only the coordination of rescue efforts and the validation of the landslide map with on-site data but also limited the map's practical use. A key lesson is that effectively managing a complex crisis demands a blend of sophisticated technical strategies and the participation of less-skilled staff who are actively involved in on-the-ground emergency response. The successful integration of these elements is crucial and necessitates thorough preparation in advance.

4.8 Conclusions

In this study, we evaluated two automated methods for rapid mapping to aid in the detection of landslides in emergency scenarios. Our evaluation took place in a selected area within the Emilia-Romagna Region (Italy), which was affected by the disaster in May 2023. The maps generated through automated processes were analyzed against a manually created map by expert geologists. From this analysis, we derived the following conclusions:

1) In our testing area, UNET outperforms the NDVI change detection method in accurately identifying landslides. The convolutional neural network approach demonstrated a lower tendency for false positives compared to NDVI, which frequently misidentifies agricultural fields as landslides. UNET's enhanced accuracy is attributed to its capability to detect landslides based on not just changes in vegetation but also the physical shape of the object.

2) All methods demonstrate their optimal performance even when trained on a limited dataset. Manually detecting landslides in roughly 2-3% of the area is enough for training the methods and achieving their highest F1-score.

3) UNET yields satisfactory outcomes solely when utilized with high-resolution images, such as 4-band RGB+NIR aerial images at a 0.2 m resolution, resampled at 2 m. Utilizing low-resolution Sentinel-2 images with 10 m pixels does not yield adequate accuracy in identifying the small landslides triggered by the event.

4) The NDVI change method exhibits better performance when evaluating the pre- and post-event vegetation state over a period by averaging multiple satellite images rather than relying on a single image. However, this approach is only feasible with Sentinel-2 images, which are freely accessible and continuously captured. Conversely, it is not viable with high-resolution images typically obtained before and after a disaster, which pertain to specific time instances.

5) Once appropriately calibrated, advanced automated techniques such as UNET have the potential to aid in landslide mapping during emergencies, particularly in identifying the affected area and determining the number of landslides. However, accurately pinpointing the extent of landslide damage—specifically, the points of impact between landslides, buildings, and roads—presents greater difficulty, as it demands very high precision, especially when dealing with numerous small landslides. Consequently, in the case of the Romagna 2023 disaster, automated mapping was unable to entirely replace manual landslide identification by experts.

6) Irrespective of whether an automated or manual approach is employed, the timely availability of high-resolution images following an event is crucial for effectively managing the crisis. Hence, existing policies governing access to high-resolution satellite data should account for the potential to share this data with public institutions engaged in emergency response efforts.

Ongoing research efforts related to the Romagna 2023 disaster entail refining UNET for broader area application, analyzing landslide susceptibility, and crafting predictive models for landslide initiation and spread. The authors extend a warm invitation to colleagues for collaborative participation, aiming to extract valuable insights from this event.

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)

4.9 References

Adriano B, Yokoya N, Miura H, Matsuoka M, Koshimura S (2020). A Semiautomatic Pixel-Object Method for Detecting Landslides Using Multitemporal ALOS-2 Intensity Images. *Remote Sensing*, 12(3), 561:1-19. https://doi.org/10.3390/rs12030561

Amatya P, Kirschbaum D, Stanley T, Tanyas H (2021). Landslide Mapping Using Object-Based Image Analysis and Open Source Tools. *Engineering Geology*, 282:1-10. https://doi:10.1016/j.enggeo.2021.106000

Behling R, Roessner S (2017). Spatiotemporal Landslide Mapper for Large Areas Using Optical Satellite Time Series Data. In: *Advancing Culture of Living with Landslides. Springer International Publishing*, Cham:143–152.

Behling R, Roessner S, Kaufmann H, Kleinschmit B (2014). Automated Spatiotemporal Landslide Mapping over Large Areas Using RapidEye Time Series Data. *Remote Sens* (Basel) 6:8026–8055. https://doi.org/10.3390/rs6098026

Bhuyan K, Meena SR, Nava L, et al (2023). Mapping landslides through a temporal lens: an insight toward multi-temporal landslide mapping using the u-net deep learning model. *GIsci Remote Sens* 60. https://doi.org/10.1080/15481603.2023

Burgess DW, Lewis P, Muller JP (1995). Topographic effects in AVHRR NDVI data. *Remote Sens. Environ*., 54(3):223–232. https://doi.org/10.1016/0034-4257(95)

Chandra N, Sawant S, Vaidya H (2023). An Efficient U-Net Model for Improved Landslide Detection from Satellite Images. PFG - Journal of Photogrammetry, *Remote Sensing and Geoinformation Science* 91:13–28. https://doi.org/10.1007/s41064-023-00232-4

Chen H, He Y, Zhang L, et al (2023). A landslide extraction method of channel attention mechanism U-Net network based on Sentinel-2A remote sensing images. *Int J Digit Earth* 16:552–577. https://doi.org/10.1080/17538947.2023.2177359

Cruden DM, Varnes DJ (1996). Landslide types and processes. In Turner, A.K., Shuster, R.L., (Eds.), Landslides: Investigation and mitigation, 247:36–75. Washington: National Academy Press.

Dou J, Chang K-T, Chen S, Yunus AP, Liu J-K, Xia H, Zhu Z (2015). Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm. *Remote Sensing*., 7(4):4318-4342. https://doi.org/10.3390/rs70404318

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

Ferrario MF, Livio F (2024). 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

Ghorbanzadeh O, Crivellari A, Ghamisi P, et al (2021). A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan). *Sci Rep* 11:14629. https://doi.org/10.1038/s41598-021-94190-9

Ghorbanzadeh O, Didehban K, Rasouli H, et al (2020). An Application of Sentinel-1, Sentinel-2, and GNSS Data for Landslide Susceptibility Mapping. *ISPRS Int J Geoinf* 9:561. https://doi.org/10.3390/ijgi9100561

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ*., 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031

Guzzetti F, Mondini AC, 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

Haneberg WC, Cole WF, Kasali G (2009). High-resolution lidar-based landslide hazard mapping and modeling, UCSF Parnassus Campus, San Francisco, USA. *Bulletin of Engineering Geology and the Environment* 68:263–276. https://doi.org/10.1007/s10064-009-0204-3

Jain N, Roy P, Martha TR, et al (2024). Causal analysis of unprecedented landslides during July 2021 in the Western Ghats of Maharashtra, India. *Landslides* 21:99–109. https://doi.org/10.1007/s10346-023-02165-w

Joubert-Boitat I, Wania A, Dalmasso S (2020). Manual for CEMS-Rapid Mapping Products. Ispra.

Keyport RN, Oommen T, Martha TR, Sajinkumar KS, Gierke JS (2018). A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinf*., 64:1-11. http://dx.doi.org/10.1016/j.jag.2017.08.015

Li Z, Shi W, Myint SW, Lu P, Wang Q (2016). Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sensing of Environment*, 175:215-230. https://doi.org/10.1016/j.rse.2016.01.003

Martinez SN, Schaefer LN, Allstadt KE, Thompson EM (2021). Evaluation of Remote Mapping Techniques for Earthquake-Triggered Landslide Inventories in an Urban Subarctic Environment: A Case Study of the 2018 Anchorage, Alaska Earthquake. *Front Earth Sci* (Lausanne) 9. https://doi.org/10.3389/feart.2021.673137

Meena SR, Soares LP, Grohmann CH, 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

Mondini AC, Guzzetti F, Reichenbach P, et al (2011). Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens Environ* 115:1743–1757. https://doi.org/10.1016/j.rse.2011.03.006

Nava L, Bhuyan K, Meena SR, et al (2022). Rapid Mapping of Landslides on SAR Data by Attention U-Net. *Remote Sens* (Basel) 14, https://doi.org/10.3390/rs14061449.

Niyokwiringirwa P, Lombardo L, Dewitte O, et al (2024). Event-based rainfall-induced landslide inventories and rainfall thresholds for Malawi. *Landslides*. https://doi.org/10.1007/s10346-023-02203-7

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* 23:2625–2648. https://doi.org/10.5194/

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

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.

Prakash N, Manconi A (2021). Rapid Mapping of Landslides Triggered by the Storm Alex, October 2020. In: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE:1808–1811.

Prakash N, Manconi A, Loew S (2020). Mapping Landslides on EO Data: Performance of Deep Learning Models vs. Traditional Machine Learning Models. *Remote Sens* (Basel) 12:346. https://doi.org/10.3390/rs12030346

Pu M, Zhao Y, Ni Z, et al (2022). Spatial-Temporal Evolution and Driving Forces of NDVI in China's Giant Panda National Park. *Int J Environ Res Public Health* 19. https://doi.org/10.3390/ijerph19116722

Qi W, Wei M, Yang W, et al (2020). Automatic mapping of landslides by the ResU-Net. *Remote Sens* (Basel) 12. https://doi.org/10.3390/rs12152487

Qu F, Qiu H, Sun H, Tang M (2021). Post-failure landslide change detection and analysis using optical satellite Sentinel-2 images. *Landslides* 18:447–455. https://doi.org/10.1007/s10346-020-01498-0

Ricci Lucchi F, Valmori E (1980). Basin-wide turbidites in Miocene, over-supplied deep-sea plain: a geometrical analysis. *Sedimentology* 27:241-270.

Ronneberger O, Fischer P, Brox T (2015). U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag:234–241.

Satriano V, Ciancia E, Filizzola C, Genzano N, Lacava T, Tramutoli V (2023). Landslides Detection and Mapping with an Advanced Multi-Temporal Satellite Optical Technique. *Remote Sens*., 15: 683. https://doi.org/10.3390/rs15030683

Shahabi H, Rahimzad M, Ghorbanzadeh O, Piralilou ST, Blaschke T, Homayouni S, Ghamisi P (2022). Rapid Mapping of Landslides from Sentinel-2 Data Using Unsupervised Deep Learning. *2022 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, Istanbul, Turkey:17-20. https://doi.org/10.1109/M2GARSS52314.2022.9840273

Soares LP, Dias HC, Grohmann CH (2020). Landslide Segmentation with U-Net: Evaluating Different Sampling Methods and Patch Sizes. *arXiv* 2020. https://doi.org/10.48550/arXiv.2007.06672

Su Z, Chow JK, Tan PS, et al (2021). Deep convolutional neural network–based pixel-wise landslide inventory mapping. *Landslides* 18:1421–1443. https://doi.org/10.1007/s10346-020-01557-6

Tanoli JI, Ningsheng C, Regmi AD, Jun L (2017). Spatial distribution analysis and susceptibility mapping of landslides triggered before and after Mw7.8 Gorkha earthquake along Upper Bhote Koshi, Nepal. *Arabian Journal of Geosciences* 10:277. https://doi.org/10.1007/s12517-017-3026-9

Tharwat A (2018). Classification assessment methods. *Applied Computing and Informatics* 17:168–192. https://doi.org/10.1016/j.aci.2018.08.003

Tucker CJ (1979). Red and Photographic Infrared linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8(2): 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

Williams JG, Rosser NJ, Kincey ME, et al (2018). Satellite-based emergency mapping using optical imagery: experience and reflections from the 2015 Nepal earthquakes. Natural Hazards and Earth System Sciences 18:185–205. https://doi.org/10.5194/nhess-18-185-2018

Wu X, Ren F, Niu R (2014). Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ Earth Sci* 71: 4725–4738. https://doi.org/10.1007/s12665-013-2863-4

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

Yang X, Zuo X, Xie W, Li Y, Guo S, Zhang H (2022). A Correction Method of NDVI Topographic Shadow Effect for Rugged Terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15. https://doi.org/10.1109/JSTARS.2022.3193419

Yang X, Chen L (2010). Using multi-temporal remote sensor imagery to detect earthquake-triggered landslides. *International Journal of Applied Earth Observation and Geoinformation* 12:487–495. https://doi.org/10.1016/j.jag.2010.05.006

Zhong C, Liu Y, Gao P, Chen W, Li H, Hou Y, Ma H. (2019). Landslide mapping with remote sensing: challenges and opportunities. *International Journal of Remote Sensing*, 41(4):1555–1581. https://doi.org/10.1080/01431161.2019.1672904

Chapter 5

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

5.1 Preface

This chapter addresses a critical challenge in disaster management: the rapid and accurate mapping of landslides following extreme weather events, such as the catastrophic rainfall that struck Emilia-Romagna in May 2023. The work represents a crucial phase of my doctoral research, aimed at enhancing real-time landslide detection through the application of advanced machine learning techniques.

The study explores the potential of two deep learning models—U-Net and SegFormer—for producing rapid, automated landslide maps with minimal training data, a necessity in emergency scenarios where time is of the essence. This research was conducted in collaboration with various institutions, and my role specifically focused on evaluating model performance and testing their applicability in complex geological conditions.

This chapter reflects both the technical challenges and the potential breakthroughs that automated approaches can offer in the context of disaster response. The insights gained here not only contribute to the development of more efficient mapping techniques but also lay the groundwork for future advancements in integrating machine learning models into emergency management systems.

The chapter is based on a current manuscript prepared for submission:

Dal Seno N.[1], Ciccarese G.[1], Evangelista D.[2], Piccolomini E.[2], Berti M.[1]. Rapid Landslide Mapping During the 2023 Emilia-Romagna Disaster: Assessing Automated Approaches with Limited Training Data for Emergency Response.

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

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

5.2 Abstract

The catastrophic rainfall events of May 2023 in the Emilia-Romagna region, Italy, triggered thousands of landslides, posing a significant challenge for emergency response. This study investigates the potential of automated landslide mapping to aid in rapid disaster response by focusing on two deep learning models: U-Net and SegFormer. Unlike traditional approaches that require extensive training data, our methodology emphasizes the necessity for quick results, given the urgent need for timely, accurate mapping in emergency scenarios.

The study was conducted in four severely affected municipalities: Casola Valsenio, Predappio, Modigliana, and Brisighella. By limiting the amount of training data to expedite the process, both U-Net and SegFormer were evaluated for their ability to deliver accurate mappings with reduced preparation time. While U-Net proved effective in identifying landslides based on changes in terrain and vegetation, SegFormer demonstrated enhanced performance in complex geological conditions, reducing false positives.

This research highlights the feasibility of integrating automated mapping techniques into emergency workflows, significantly shortening response times while maintaining a high level of accuracy. Although manual verification remains necessary for refining results, particularly in complex situations, this study demonstrates that with minimal training data, automated mapping can play a crucial role in emergency management, improving the speed and efficiency of disaster response.

5.3. Introduction

Rapid and accurate landslide mapping over large areas is a challenging task, particularly in the context of emergency management following regional-scale rainfall or earthquake events (Holbling et al. 2017; Iverson et al. 2015; Casagli et al. 2016, Dal Seno et al. 2024). The ability to quickly and accurately identify landslides is crucial for effective disaster response and mitigation efforts. This is especially true when a large number of landslides must be mapped to support emergency management. In such cases, remote sensing and machine learning approaches can potentially improve rapidity in mapping (Ferrario et al. 2024; Guzzetti et al., 2012; Sameen and Pradhan 2019, Chen et al. 2018, Ye et al. 2019, Tang et al. 2022, Ji 2020). The challenges in landslide mapping are multifaceted, depending on factors such as immediate access to high-resolution, cloud-free imagery, deployment of skilled analysts, effectiveness of image analysis software, and difficulties in identifying certain landslides due to various environmental factors like small size, tree cover, or minimal impact on terrain and land cover.

Current solutions for emergency landslide mapping scenarios vary in their approaches and effectiveness. In Europe, the Copernicus Emergency Management Service (CEMS) utilizes mid- to high-resolution satellite imagery for rapid mapping of landslides and flooded areas

(Joubert-Boitat et al., 2020). However, these services often face limitations in accuracy and accessibility. Only a limited portion of the imagery Copernicus uses is accessible to third parties for further examination or for running automated landslide detection algorithms. While datasets from Sentinel-2 are freely distributed (Amatya et al., 2023), acquiring higher resolution datasets, like those from Pléiades 1A/B, requires specific orders and faces distribution restrictions. This limitation significantly complicates the timely access to these datasets, potentially rendering them insufficient for practical emergency response.

Automated landslide recognition techniques using convolutional neural networks (CNN) algorithms are emerging as promising alternatives to manual methods (Holbling et al. 2017). Recent studies have demonstrated their effectiveness in real-world emergency scenarios. 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. 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 contexts. Building on this, Prakash and Manconi (2021) demonstrated the practical utility of these methods by rapidly mapping landslides triggered by a severe storm event. These studies collectively highlight the potential of automated methods to provide quick and accurate results in emergency situations, crucial for timely disaster response and management. The focus on real-world applicability in these works underscores the growing importance of such technologies in enhancing practical disaster management capabilities, particularly in scenarios requiring rapid response over large or complex terrains.

The May 2023 events in the Emilia-Romagna Region of Italy provide a compelling context for this study. Two intense cyclonic events, occurring from May 1-3 and May 16-18, resulted in up to 500 mm of rainfall in the Romagna Apennines, triggering approximately 70,000 landslides and widespread flooding. The disaster caused extensive damage across the provinces of Forlì-Cesena, Ravenna, Bologna, Modena, and Rimini. Public property saw over 4,000 incidents of damage, agricultural losses impacted 2,900 farms, and essential infrastructure, including roads and bridges, was severely affected, isolating many locations. The total area impacted by landslides exceeded 70 square kilometers, emphasizing the event's vast scale. The flooding alone inundated 540 square kilometers of land. The aftermath included damages worth several billion Euros and 17 fatalities, highlighting the crisis's severe human and economic toll. This disaster underscores the urgent need for efficient and accurate landslide mapping techniques to support rapid response and mitigation efforts in future emergencies (Berti et al 2024a – “Emergency mapping: lessons learned from 2023 landslide event in Romagna”, under review; Technical-Scientific Commission Report, 2023).

This study leverages the ground-truth dataset collected during the Emilia-Romagna crisis to evaluate the benefits and limitations of automated mapping algorithms. Specifically, we assess the effectiveness of U-Net and SegFormer algorithms, focusing on their practical use in post-

event emergencies. We examine the requirements for ground-truth data to achieve reliable outcomes and the impact of image resolution and quality on the results. By addressing these aspects, we seek to determine if automated approaches can offer significant benefits over conventional, manually supervised mapping, especially in rapidly responding to widespread landslide disasters. This research is crucial for developing more efficient and timely landslide mapping techniques that can support critical decision-making in future emergency situations, bridging the gap between academic advancements and practical, on-the-ground applications in crisis management.

5.4. The Romagna May 2023 Rainfall Events

5.4.1 Rainfall Events

The Emilia-Romagna region experienced unprecedented meteorological events in May 2023, detailed extensively by Berti (Berti et al., 2024b - " Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy", under review) and Foraci et al. (2023). Two major rainfall episodes, from May 1-3 and May 16-17, led to severe floods and landslides, surpassing the damage recorded during the historic floods of 1939 and 1966. During the first event (May 1-3), a stationary low-pressure system caused extraordinary rainfall, with cumulative rainfall peaking at 274.4 mm in Le Taverne (Modigliana, Forlì-Cesena). The second event, from May 16 to 17, was similarly severe. Another low-pressure system, combined with bora winds and moist air from the Mediterranean, resulted in even higher rainfall accumulations. These unprecedented amounts exceeded historical maxima, with return periods estimated to be over 500 years for the maximum areal rainfall in May 2023 for 4 out of the 7 sample basins used in the *Report of the Technical-Scientific Commission established by Regional Council Resolution No. 984/202* (Technical-Scientific Commission Report, 2023), aimed at analyzing the extreme meteorological events of May 2023.

The cumulative precipitation from May 1 to 17 averaged 221.40 mm regionally, with some locations reporting up to 609.8 mm, such as Trebbio (Modigliana, Forlì-Cesena), Monte Albano (Casola Valsenio, Ravenna), San Cassiano (Brisighella, Ravenna), and Predappio (Forlì-Cesena). These values represent one-third to one-half of the annual climatic precipitation, underlining the exceptional nature of this period. Figure 5.1b illustrates the cumulative precipitation distribution from May 1 to 17 across these communes.

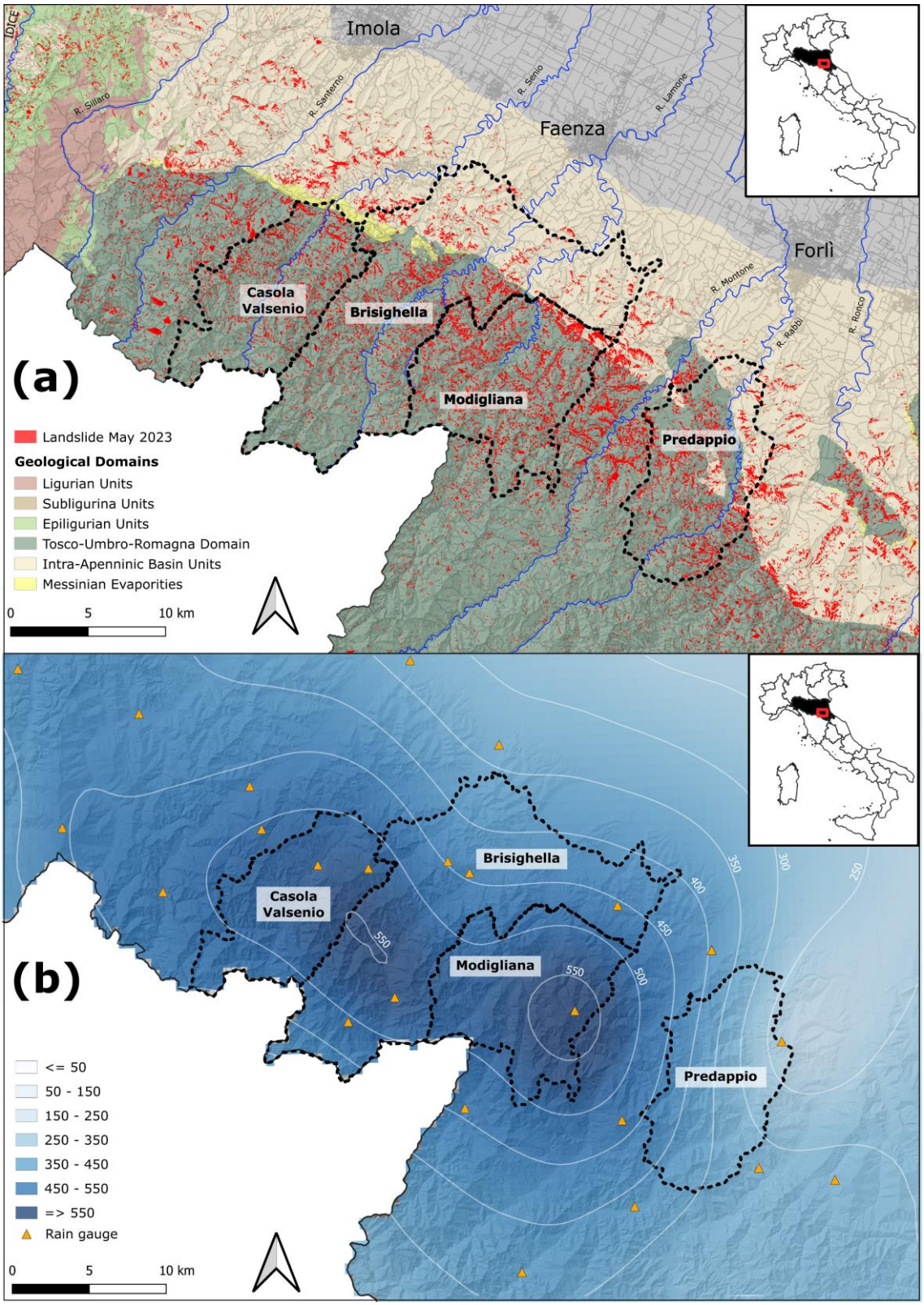


Figure 5.1. Overview of the Effects of the Extreme Rainfall Event in Romagna, May 2023. a) Manually mapped landslide distribution across the study areas. b) Cumulative precipitation map from May 1 to May 17, 2023.

5.4.2 Types of Triggered Landslides

5.4.2.1 Rapid Earth and Debris Slides

A significant portion of the landslides were rapid earth (ES) and debris slides (DS) (Berti et al., 2024b). These movements, occurring on steep, inclined slopes, especially near watercourses and road embankments, formed large, denuded surfaces by removing vegetation and depositing reworked material at the base of slopes. These slides, characterized by the movement of surface debris composed mainly of sand, silt, and the upper soil layer including vegetation, often evolved into flow-like movements if the topography allowed. Unlike classic slides, these materials did not move as a cohesive block but flowed along the slope, creating accumulations of completely reworked material. Such events were common on steep, forested slopes and often impacted mid-slope roads (Fig. 5.2b) (Technical-Scientific Commission Report, 2023).

5.4.2.2 Earth and Debris Flows

Additionally, numerous earth (EF) and debris flows (DF) occurred (Berti et al., 2024b). These flows, both channeled and unchanneled, were characterized by their ability to travel over planar slopes or within gullies, often originating from single points. On steeper slopes, they exhibited typical debris flow morphology with a source area, a narrow channel, and a lobate deposit zone. These flows could travel hundreds of meters, carrying mud, debris, and uprooted vegetation, causing extensive damage to communication routes and buildings (Fig. 5.2c, 5.2d). Unchanneled flows were characterized by straight, constant-width paths, whereas channeled flows had more irregular shapes and could form new channels through the mobilization of existing deposits.

5.4.2.3 Rock Slides Along Bedding Planes

Rock slides (SCST) along bedding planes were also significant (Berti et al., 2024b). These slides involved the rock mass sliding along stratification planes, marked by perimeter fractures. These types of landslides typically occurred on less steep, dip slopes and often involved large areas, sometimes exceeding 10 hectares. Although their overall affected area often required interpretative mapping, they contributed significantly to the widespread landscape changes observed (Fig. 5.2a).

The May 2023 landslides caused extensive landscape alterations, stripping vast areas of vegetation and soil cover. The rapid movement of debris and rock masses damaged numerous roads across the Romagna region, affecting detachment zones, flow tracks, and deposition areas. The severity of these events led to the destruction of houses and fatalities, prompting a state of emergency. Many individuals were forced to evacuate, and access to several villages

and valleys was hindered due to the extensive road and infrastructure damage (Pizziolo et al., 2024, Berti et al., 2024b).

5.4.3 Motivation

High-precision manual landslide mapping, while accurate, requires substantial time—a luxury often unavailable in emergency situations. The May 2023 event in Emilia-Romagna has underscored the need to develop more efficient approaches for future emergencies. This study is motivated by the critical question: How can we more effectively address landslide mapping in time-sensitive scenarios?

Our research aims to simulate an emergency situation similar to the aftermath of the May 2023 disaster. In such scenarios, a post-event imagery must be available, but comprehensive mapping of all affected areas is not immediately feasible. We explore whether mapping landslides manually in a single municipality—in this case, Casola Valsenio—can provide sufficient training data to automatically map landslides in other affected communities.

The core motivation of this study is to determine if a model trained on data from one municipality can effectively map landslides in neighboring areas. We seek to understand the quality and accuracy of such automated mapping and to identify the geographical and geological limits of its applicability.

To this end, we apply our trained model to the municipalities of Predappio, Modigliana, and Brisighella. These areas largely share similar lithology and landslide typology with Casola Valsenio, with the exception of a portion of Brisighella situated in Pliocene blue clays (FAA). This approach allows us to assess the model's performance across similar geological contexts while also testing its limits in slightly divergent conditions.

By addressing these questions, we aim to develop insights that could significantly improve the efficiency and effectiveness of landslide mapping in future emergency situations, potentially saving crucial time in critical response periods.

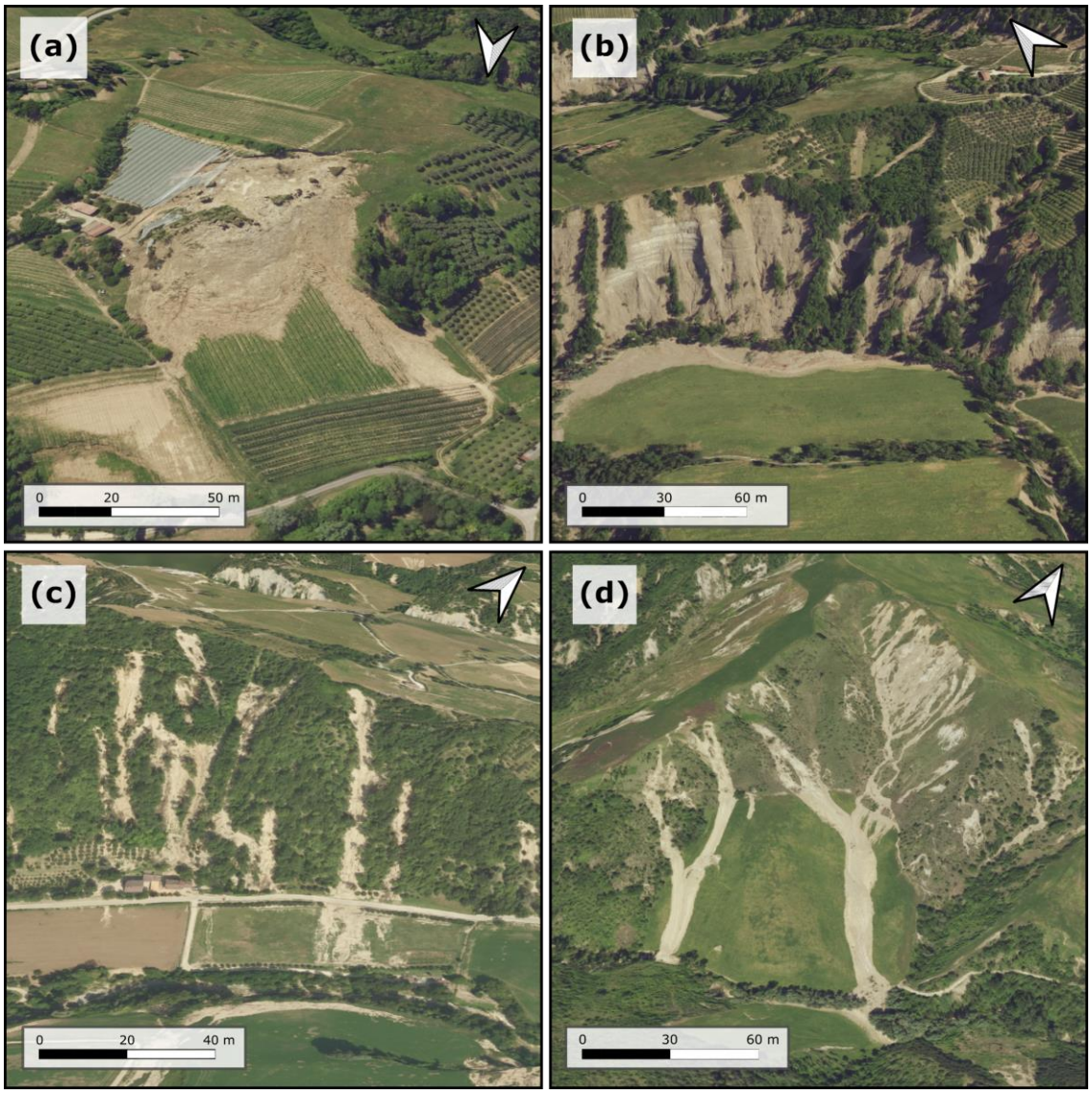


Figure 5.2. Examples of Landslides in Romagna from May 2023 event.

a) Rockslide (SCST) at Baccagnano, Brisighella, triggered during the night of May 2-3 (WGS84: 44.207, 11.774).

b) Shallow debris slide (DS) at Monte Albano, Casola Valsenio (WGS84: 44.236, 11.668).

c) Channeled debris flow (DF) at Modigliana (WGS84: 44.148, 11.774).

d) Unchanneled debris flow (DF) at Predappio (WGS84: 44.118, 11.968).

Images sourced from the Geoportale 3D Regione Emilia-Romagna.

5.5. Methods

5.5.1 Manual landslide mapping

A detailed inventory of landslides was manually compiled using high-resolution aerial imagery from May 2023 (0.2 m pixel) and 2020 (0.2 m pixel). Geologists, who had inspected the landslide sites shortly after the event, conducted this work, leveraging their field experience to

accurately identify landslide features in the photos (Berti et al., 2024b). This process included visual inspection and digitization within a GIS framework, using scales from 1:1000 to 1:200. A 1x1 km reference grid was utilized to aid surveyors and monitor progress. NDVI change maps from Sentinel-2 data, NIR band imagery, hillshade renderings from a 5 m resolution DEM, and topographic maps supported the analysis. Additionally, a 3D visualization tool was used for dynamic perspectives.

Within the GIS, landslide outlines (polygonal layers) and initiation points (point layers) were digitized. Only landslides from May 2023 were included; those unchanged since 2020 were excluded. For debris flows with multiple source areas, several initiation points were recorded, typically in the highest parts of the detachment zones. The landslide layer's attribute table used codes to differentiate types such as rotational debris slide (RDS), translational debris slide (DS), debris flow (DF), rock slide (RS), earth slide (ES), and earth flow (EF). An inventory of landslide-induced damages was also created, marking affected roads and buildings, with attributes distinguishing their location and proximity to a 30 m buffer zone around the landslide outline (Berti et al., 2024b).

Most May 2023 landslides were first-time failures, easily identifiable in aerial images. Debris flows and slides cleared vegetation, revealing bare soil in deposition zones, while translational rock block slides caused significant landscape changes. Some challenges included small-scale events and those obscured by canopy cover. Despite these challenges, the inventory is estimated to be over 90% complete.

5.5.2 Study area

5.5.2.1 Casola Valsenio

The automated mapping techniques were evaluated in Casola Valsenio (Ravenna province), a region heavily impacted by the severe hydrogeological events of May 2023. Spanning 84 km², Casola Valsenio is dominated by the Marnoso-Arenacea Formation (FMA), which is stratified into four members based on the ratio of sandstone to pelites (Fig. 3a). The Senio River runs through the area, intersected by smaller tributaries at sharp angles. The landscape's asymmetry, with slopes ranging from 30° to 70°, is characteristic of the Marnoso-Arenacea Formation. Prior to May 2023, the Geological Survey had recorded 730 ancient, inactive landslides, mostly rock-block slides on slopes aligned with bedding planes. After the disaster, this number surged to 5.617 new landslides, transforming the terrain. With a population of 2.650, Casola Valsenio serves as a key connection between the Po Plain and Tuscany. The disaster disrupted local infrastructure, blocking the main road and isolating the southern region for weeks, with nearly 70 homes and 200 roadways severely impacted.

5.5.2.2 Predappio

In Predappio (Forlì-Cesena province), another focus area for the study, the effects of the May 2023 disaster were profound. This municipality, covering 91 km², is largely composed of the Marnoso-Arenacea Formation and the Blue Clays Formation (FAA) in the higher regions (Fig. 3b). The Rabbi River and its tributaries carve through the terrain, characterized by steep, asymmetrical slopes ranging from 25° to 65°. Prior to the catastrophic event, 840 landslides had been recorded, primarily involving rock-block slides and debris flows. Following the disaster, this number escalated to 6.840, altering the landscape significantly. Predappio, home to 6.300 residents, plays a critical role in linking the Po Plain with central Italy. The region experienced weeks of isolation due to a major landslide blocking the main road, impacting 19 homes and damaging 138 roadways, with long-lasting consequences.

5.5.2.3 Modigliana

Modigliana, another municipality in the Forlì-Cesena province, was severely affected by the May 2023 hydrogeological disaster. This area spans 101 km² and is primarily composed of the Marnoso-Arenacea Formation, with a portion of Blue Clays in the northeast (Fig. 3c). The Tramazzo River flows through the region, with several tributaries joining it at right angles. The slopes in Modigliana feature steep inclinations between 35° and 70°, shaped by the orientation of the rock layers. Before the disaster, 795 landslides were documented, mostly complex and rock-block slides. After the May 2023 event, the number of landslides jumped to 6.862, drastically reshaping the landscape. With 4,500 residents, Modigliana serves as a crucial connector between the Po Plain and Tuscany. The disaster blocked the main access road for weeks, disrupting daily life and affecting 51 homes and 155 roadways, leaving the community to face long-term recovery challenges.

5.5.2.4 Brisighella

Brisighella, located in the Ravenna province, covers 194 km² and experienced significant disruption from the May 2023 disaster. The municipality is primarily composed of the Marnoso-Arenacea Formation, while the northern section, covering about 75 km², is dominated by the Blue Clays Formation, separated by a small zone of gypsum (Fig. 3d). The Lamone River bisects the area, with numerous tributaries intersecting at right angles. The landscape features steep, asymmetrical slopes ranging from 30° to 70°. Prior to the event, the Geological Survey had recorded 2.100 landslides, mainly ancient, inactive rock-block and complex slides. Post-disaster, an additional 6.274 landslides were recorded, exacerbating the region's vulnerability. Brisighella, with a population of 7.600, faced severe disruptions, with landslides blocking key

access routes and affecting 66 homes and 261 roadways. The recovery process is ongoing as the community grapples with the extensive damage.

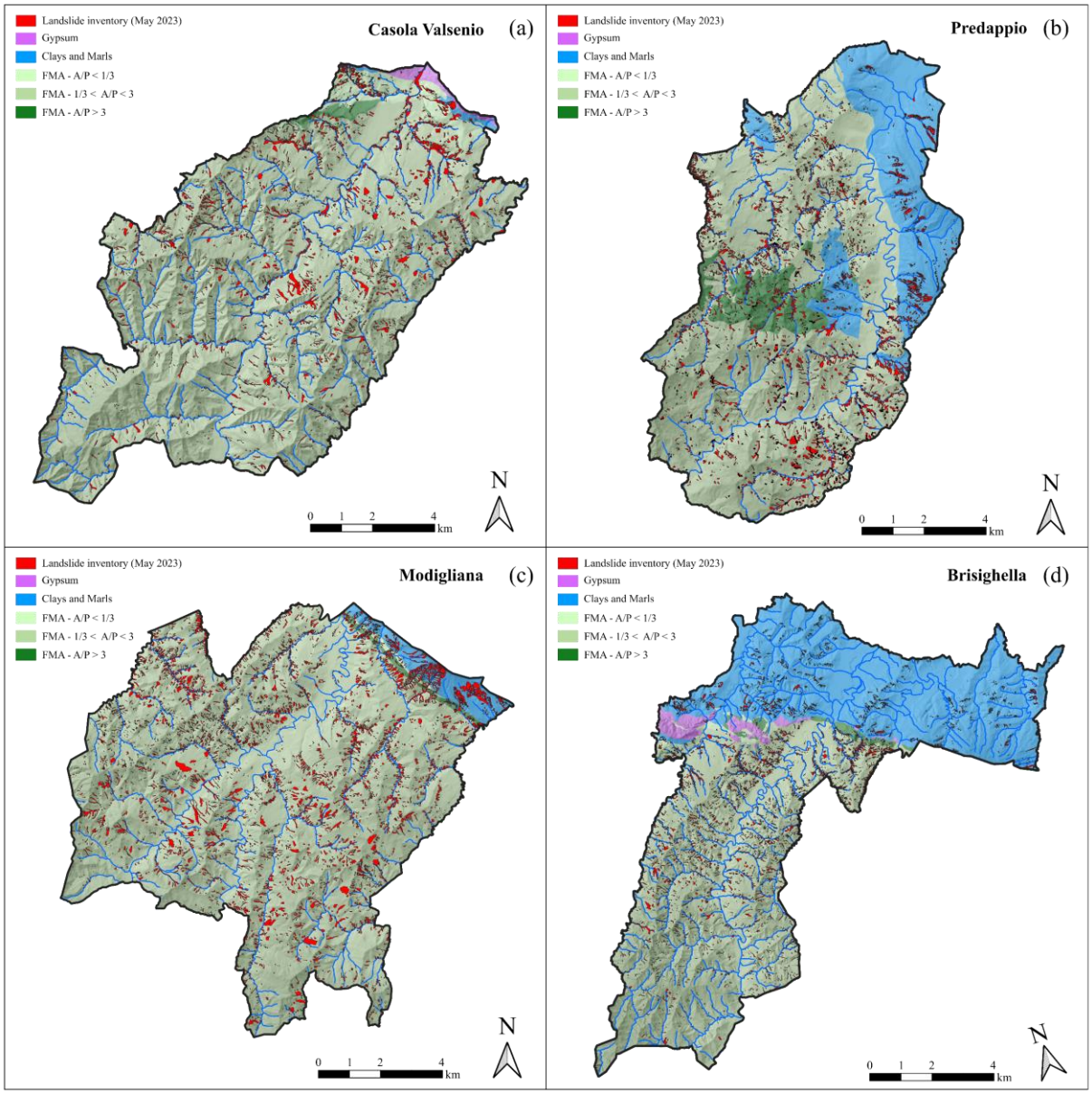


Figure 5.3. Schematic geological map of the municipalities of Casola Valsenio (**a**), Predappio (**b**), Modigliana (**c**) and Brisighella (**d**). FMA = Marnoso Arenacea Formation, categorized into three distinct members based on the prevailing Arenites/Pelites (A/P) ratio.

5.5.3 Dataset

Our model training utilized a multi-layer input approach, combining various high-resolution imagery and derived products to capture the complex characteristics of landslides. All layers were resampled to a 2-meter resolution to ensure computational efficiency and pixel-to-pixel correspondence across datasets. Additionally, all layers were normalized to a float32 range of

0 to 1, as this normalization often improves the performance of machine learning models by ensuring consistency in the input data scale.

The following layers were included:

- CGR Imagery (Fig. 5.4b): High-resolution aerial photographs (0.2 m original resolution) commissioned by the Emilia Romagna Region, captured on May 23, 2023, shortly after the landslide events. These images comprise 4 bands (RGB + NIR) and provide detailed post-event landscape information.
- Slope (Fig. 5.4d): Derived from the 5x5 m Digital Terrain Model (DTM) provided by the Emilia Romagna Region's geoportal. This DTM, originally based on 1:5000 scale Regional Technical Map and updated with 2009 LiDAR data, offers crucial topographic information for landslide susceptibility analysis.
- AGEA Imagery (Fig. 5.4a): High-resolution (0.2 m) aerial photographs from 2020, captured by the Italian Agency for Agricultural Payments. These 4-band (RGB+NIR) images represent pre-event conditions and are valuable for land use and vegetation change analysis.
- NDVI Change Sentinel-2 Map (Fig. 5.4f): Derived from Sentinel-2 satellite data (10 m resolution). Berti et al., 2024a (" Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy", under review) utilized multiple datasets to capture both pre- and post-event conditions:
    - S1: Multiple images from May and June 2022
    - S2: One image taken exactly a year before the event
    - S3: First post-event image with minimal cloud cover (May 23, 2023)
    - S4: Eight additional low-cloud-coverage images captured over the following month.
- NDVI Change CGR Map (Fig. 5.4e): Derived from high-resolution (0.2 m) aerial imagery. Berti et al., 2024a (" Automated Mapping During an Emergency: Lessons Learned from the 2023 Landslide Event in Romagna, Italy", under review), resampled the original 0.2-meter resolution images to 1-meter resolution using bilinear interpolation, applying this process to the original spectral bands to preserve data integrity.
- Landslides map (Fig. 5.4c): A binary raster derived from the manual landslide mapping process. This inventory was compiled by experienced geologists using high-resolution aerial imagery from May 2023 and 2020 as described in section 5.5.1.

By integrating these diverse data layers, we aim to provide our models with a rich, multi-dimensional representation of the landscape, enabling more accurate and robust landslide detection and delineation.

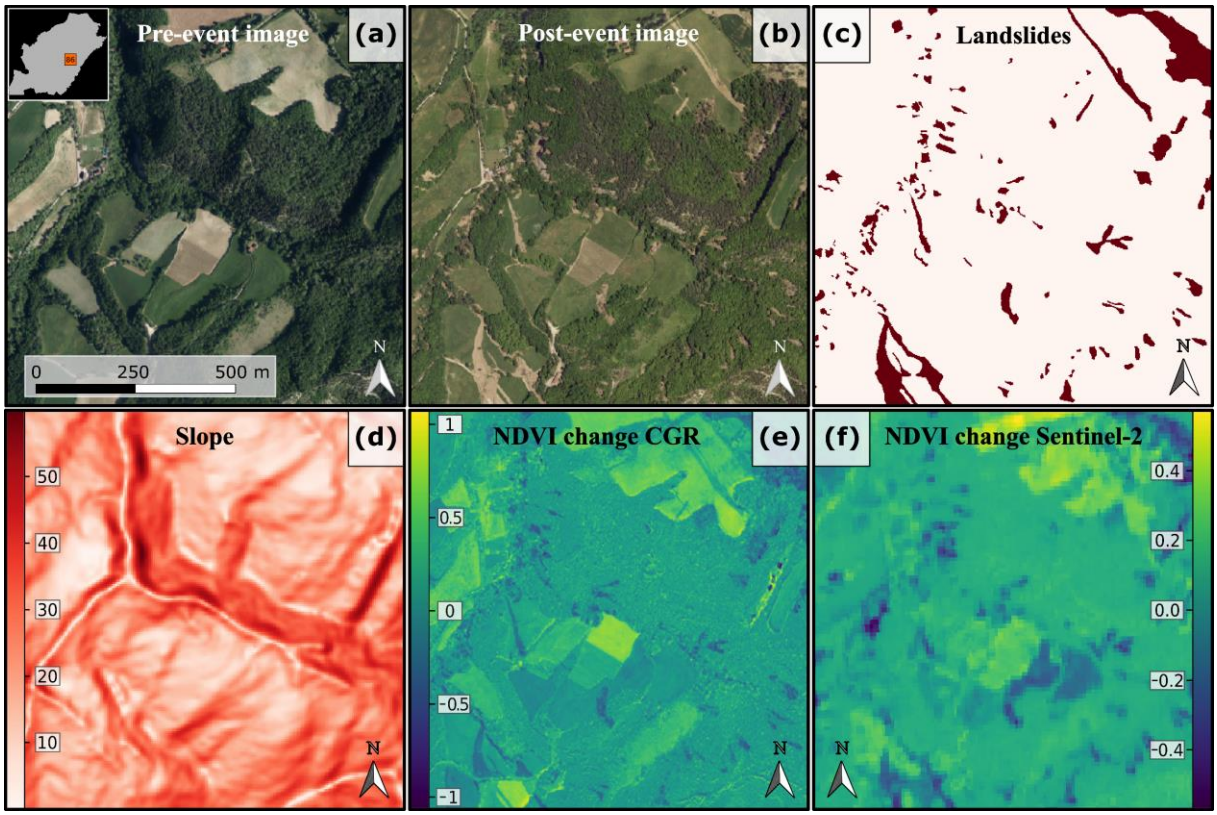


Figure 5.4. Example of layers included in tile no. 86. **a**) AGEA imagery 2020, pre-event in RBG **b**) CGR aerial image 2023, post event; **c**) Binary raster of manually mapped landslides **d**) Slope map; **e**) NDVI Change CGR Map **f**) NDVI Change Sentinel-2.

5.5.4 Deep Learning Semantic Segmentation Models

5.5.4.1 U-Net

UNet 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 upsamples 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. UNet'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 UNet 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 1000 epochs.

To evaluate the performance of the model during training, we monitored the F1 score and Intersection over Union (IoU) metrics. The F1 score provides a balance between precision and recall, while IoU measures the overlap between predicted and actual landslide areas, giving a comprehensive evaluation of the model's segmentation accuracy. Additionally, we implemented an early stopping callback to halt training if the validation loss did not improve for 50 epochs (Prechelt, 1998).

5.5.4.2 SegFormer

We implemented SegFormer, a state-of-the-art architecture for semantic segmentation of multispectral landslide imagery, chosen for its efficiency, flexibility in handling variable-resolution imagery, and robust performance on multispectral datasets. SegFormer, introduced by Xie et al. (2021), is designed to overcome limitations of previous vision Transformer models, offering both high efficiency and strong performance. While the original SegFormer architecture was designed for 3-channel RGB images, the SegformerForSemanticSegmentation implementation from Hugging Face Transformers library (Wolf et al., 2020) we used 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:

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 1000 epochs.

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

5.5.5 Training-Testing Split

While conventional practice often dictates a 70/30 or 80/20 ratio for training and testing datasets (Hastie et al., 2009), this study adopts an innovative approach to simulate real-world emergency mapping conditions, as exemplified by the May 2023 event. The model was exclusively trained on data from the Casola Valsenio municipality and subsequently applied and evaluated on three additional municipalities: Predappio, Modigliana, and Brisighella. This methodology emulates an emergency scenario where, following the mapping of one municipality, rapid and automated mapping of subsequent areas can be achieved, albeit with inherent limitations.

To facilitate analysis, the study area was segmented into 1x1 km grid tiles, as illustrated in Fig. 5.5. Each 1 km² tile corresponds to 496x496 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, a stratified random sampling approach was employed. From the total 97 tiles in Casola Valsenio, 70% (68 tiles) were randomly selected for training, with the remaining 30% (29 tiles) reserved for initial testing and validation. This approach ensures a representative sample while preserving the integrity of the testing data (Fig. 5.5).

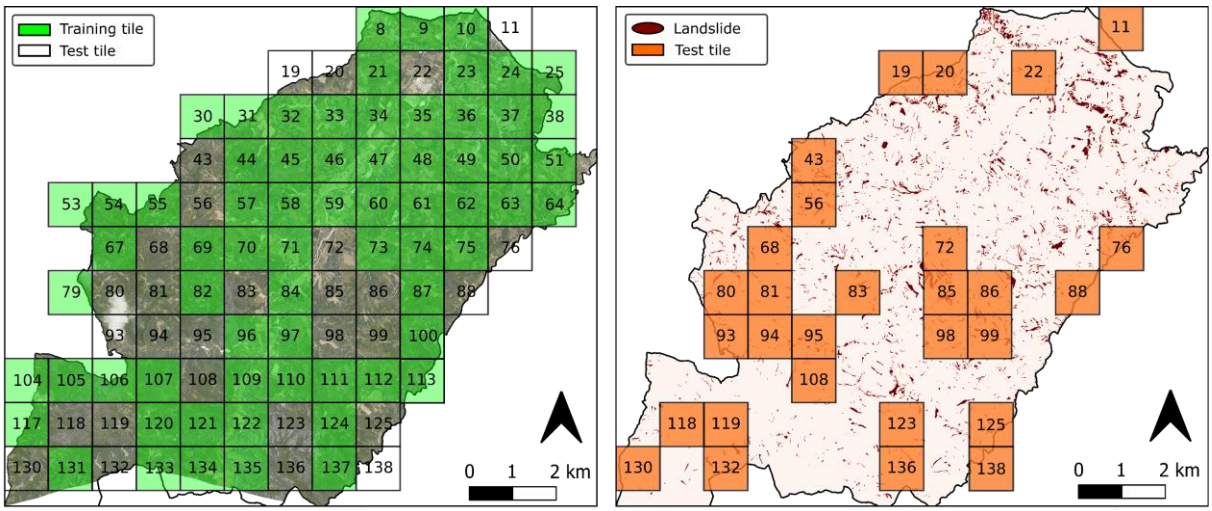


Figure 5.5. Training and test tiles used for models' construction in Casola Valsenio, illustrating the spatial split of the area.

The U-Net and SegForm models were trained using various combinations of the dataset as illustrated in Table 11. Each combination was designed to correspond to different emergency scenarios, acknowledging that comprehensive data acquisition is not always feasible in post-event situations. The cases are as follows:

- Case 1 (U1 and S1): This represents the baseline scenario, simulating the most limited data availability typical of the immediate aftermath of an event. In this case, only the post-event imagery and slope data are accessible. This scenario reflects the earliest stage of emergency response when rapid aerial imagery has been acquired, but pre-event comparison data is not yet available.
- Case 2 (U2 and S2): This case builds upon the baseline by incorporating publicly available satellite data. In addition to the post-event imagery and slope data, it includes NDVI change information derived from Sentinel-2.
- Case 3 (U3 and S3): This scenario introduces the availability of pre-event imagery. It includes pre- and post-event images along with slope data.
- Case 4 (U4 and S4): This case represents a more comprehensive data set, incorporating NDVI change information. It includes pre- and post-event imagery, slope data, and NDVI change calculated from the high-resolution images
- Case 5 (U5 and S5): This final case represents the ideal scenario with access to all possible data layers. It includes CGR pre- and post-event imagery, slope data, NDVI change from high resolution imagery, and NDVI change from Sentinel-2. This comprehensive dataset allows for the most thorough analysis, combining high-resolution aerial data with broader satellite-derived information.

This progression from Case 1 to Case 5 simulates the evolution of data availability in a post-event scenario, ranging from the most limited to the most comprehensive. It enables the evaluation of U-Net and SegForm models under varying conditions of data accessibility, reflecting realistic situations that may occur during emergency management and response efforts.

Table 5.1. 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	CGR Post-event	Slope	AGEA Pre-event	ΔNDVI-CGR	ΔNDVI-S2
U1	U-Net	✓	✓	X	X	X
S1	SegForm	✓	✓	X	X	X
U2	U-Net	✓	✓	X	X	✓
S2	SegForm	✓	✓	X	X	✓
U3	U-Net	✓	✓	✓	X	X
S3	SegForm	✓	✓	✓	X	X
U4	U-Net	✓	✓	✓	✓	X
S4	SegForm	✓	✓	✓	✓	X
U5	U-Net	✓	✓	✓	✓	✓
S5	SegForm	✓	✓	✓	✓	✓

5.5.6 Model Application

Following the training phase, we applied the models to the entirety of Predappio, Modigliana, and two distinct geological zones within Brisighella (FMA and FAA). Crucially, while ground truth (Fig. 5.1a) data was available for these municipalities, we deliberately chose not to use it

for model training. Instead, this data was reserved exclusively for evaluating the performance of our models.

A significant limitation arises from the predominance of the Marnoso-Arenacea Formation (FMA) in the training dataset from Casola Valsenio (Fig. 5.3a). To address this geological bias, we isolated the portion of the Brisighella municipality situated in the Pliocene Blue Clays Formation (FAA) (Fig. 5.3d).

This area is predominantly characterized by earth flows (EF) (Berti et al., 2024a, 2024b), which exhibit fundamentally different dynamics due to the unique lithological, morphological, and vegetative factors compared to those present in other municipalities (Fig. 5.6ab).

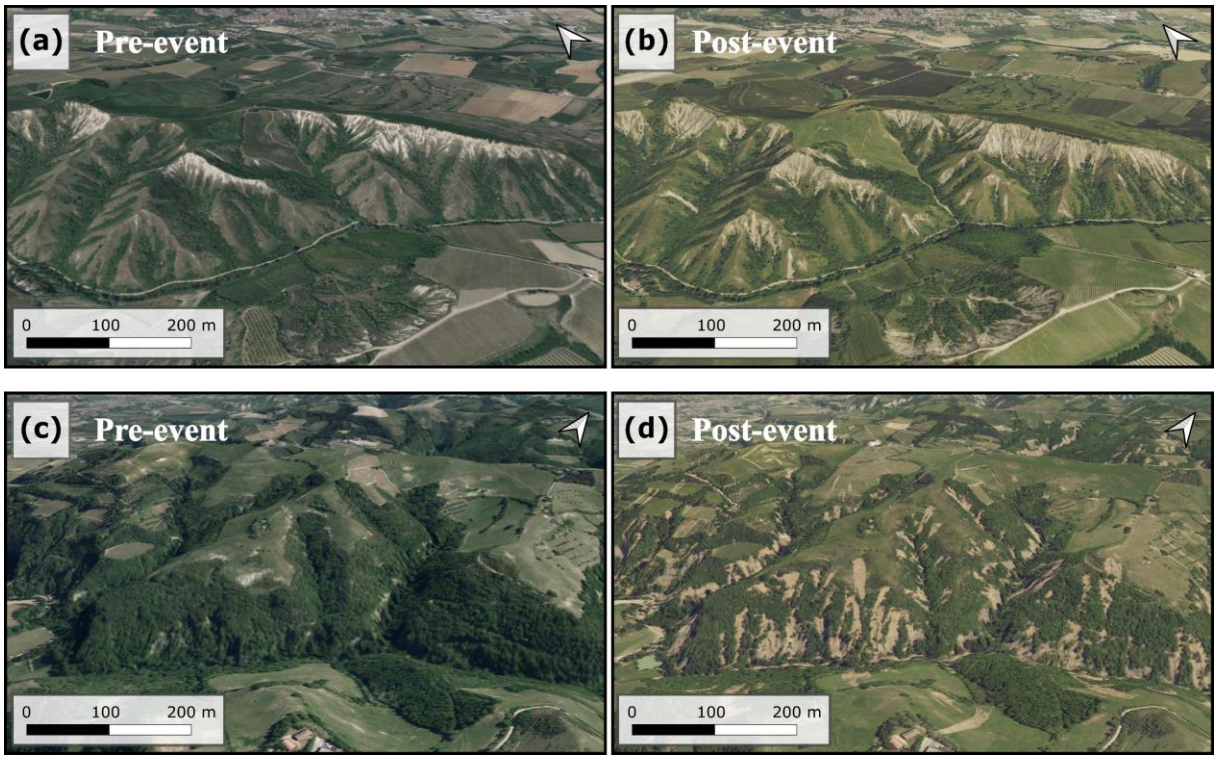


Figure 5.6. Comparison examples are shown between pre-event (**a**) and post-event (**b**) earth flow (EF) landslides in the Blue Clays Formation (FAA) and pre-event (**c**) and post-event (**d**) debris slide (DS) in the Marnoso-Arenacea Formation (FMA), both triggered by the May 2023 disaster in Brisighella.

5.5.7 Evaluation Metrics and Expert Judgment

5.5.7.1 Metrics

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 is particularly useful in imbalanced classification problems, such as landslide detection, where the positive class (landslide) is typically much smaller than the negative class (non-landslide). The F1 score is calculated as:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad [5.1]$$

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 crucial 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:

$$IoU = \frac{TP}{TP + FP + FN} \quad [5.2]$$

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.

The choice to calculate these metrics on the entire test images, rather than averaging across cells, allows us to capture the models' performance on landslides that may span multiple grid cells and provides a more realistic assessment of the models' capabilities in practical applications.

5.5.7.2 Expert Judgment

In landslide mapping, particularly during emergency phases, a higher algorithmic score does not necessarily correspond to a higher quality map. Algorithms may find preferential paths to increase scores at the expense of the specific quality required for practical application. While this issue is typically addressed by selecting the most appropriate score based on which components of the Confusion Matrix we wish to emphasize, in emergency automated mapping, not all components of the confusion matrix carry equal importance.

To address this challenge, two of the authors in this study provided their expert judgment on the quality of the automatically generated on Modigliana's 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. Their subjective assessments primarily focused on:

- Overall mapping quality.
- False positives in agricultural fields, where changes in plowing conditions between pre- and post-event images can be misinterpreted as landslides.

- False positives on buildings and roads.
- False positives along river margins, often mistaken for landslides due to flooding-induced color changes that may appear similar to the algorithm.
- The shape and delineation of the created landslide polygons.

This expert evaluation was conducted with the practical perspective of which mappings would be most useful in an emergency scenario. For instance, it is generally faster to delete an erroneously created landslide polygon than to identify and map a missed landslide. This approach aims to balance algorithmic performance with real-world applicability, ensuring that the automated mapping results are not only statistically sound but also practically valuable in emergency response situations.

5.6 Results

The machine learning models were trained using 68 tiles (1x1 km) within the Casola Valsenio municipality and applied to map landslides in Predappio, Modigliana, Brisighella MA, and Brisighella FAA. Fifty result maps were generated using different combinations of input layers (Table 5.1), with the resulting scores summarized in Table 5.2.

Comparing the scores (F1 and IoU in Table 5.2) against expert evaluation revealed a strong correspondence between quantitative metrics and qualitative assessments. Maps with higher scores were generally evaluated more favorably by experts, validating the reliability of our quantitative metrics. Notably, the range of scores produced by SegForm was narrower than UNet, suggesting SegForm may be less influenced by input parameter variations.

Both UNet and SegForm performed well in models utilizing multiple input layers (Table 5.2). Interestingly, SegForm showed good performance even with only three layers. This suggests that an advanced state-of-the-art model like SegForm might achieve satisfactory results using only pre- and post-event imagery, potentially without requiring NDVI change data.

The importance of pre-event imagery was highlighted by the performance drop observed in model S2, which lacked this input. Our findings indicate that, in our case, the absence of pre-event imagery cannot be adequately compensated by lower-resolution NDVI change data (e.g., from Sentinel). The coarser resolution appears insufficient to bridge this informational gap.

Analysis of Table 5.2, where the best and worst scores are highlighted in green and red respectively, reveals varying performance between UNet and SegForm models. For instance, models U2 and S2, trained using post-event imagery, slope, and pre-event imagery (Table 11), yielded contrasting results. S2 achieved the highest F1 and IoU scores for Predappio, Brisighella MA, and FAA, while its U2 counterpart obtained lower scores.

Table 5.2. Quality of the automatic mapping, based on F1 and IoU scores, achieved by UNet and SegFormer algorithms with different combinations of input layers, detailed in Table 11.

Models' Name	Casola Valsenio		Predappio		Modigliana		Brisighella MA		Brisighella FAA	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
U1	0.67	0.51	0.40	0.27	0.39	0.26	0.46	0.32	0.25	0.15
S1	0.61	0.44	0.48	0.31	0.46	0.30	0.54	0.37	0.35	0.21
U2	0.68	0.52	0.38	0.26	0.43	0.29	0.50	0.35	0.21	0.13
S2	0.60	0.42	0.49	0.32	0.47	0.31	0.53	0.36	0.34	0.20
U3	0.59	0.43	0.32	0.21	0.36	0.24	0.40	0.26	0.30	0.13
S3	0.62	0.45	0.54	0.37	0.54	0.37	0.58	0.41	0.40	0.25
U4	0.68	0.52	0.52	0.36	0.57	0.40	0.56	0.40	0.38	0.26
S4	0.67	0.51	0.52	0.35	0.53	0.36	0.57	0.40	0.39	0.24
U5	0.66	0.50	0.49	0.34	0.52	0.37	0.53	0.38	0.39	0.27
S5	0.61	0.44	0.53	0.36	0.53	0.36	0.56	0.39	0.40	0.25

However, given the similarity of scores across different municipalities, small differences (e.g., 0.01) in F1 score and/or IoU may not necessarily indicate significantly higher quality mapping. Examining the results of models U3 and S3, trained using pre- and post-event imagery, slope, and NDVI from CGR (Table 5.1), both models demonstrate good performance. Figure 5.7 illustrates the mapping results of U3 and S3 models for various municipalities, showing good and similar mapping capabilities, particularly for areas extensively affected by Debris Slides (DS) that exhibit similar characteristics in shape and vegetation removal, resulting in a distinct change in coloration.

Figure 5.8 highlights example situations where the same mappings (U3 and S3) encounter difficulties, leading to errors:

- In Casola Valsenio, the S3 model, likely confused by the different shape of the rock slide on layer (SCST) compared to the more common Debris Slides (DS) on which it was trained, incorrectly maps this landslide by dividing it into multiple smaller portions, resulting in False Negatives.

- In Predappio, both U3 and S3 models miss a series of landslides located on a slope that was in shadow at the time of the post-event image capture. This difference in coloration misleads the model, which fails to map any landslides on this slope, resulting in False Negatives.
- In Modigliana and Brisighella MA, both models tend to map smaller or larger portions of some cultivated fields that frequently change appearance due to plowing and cyclic vegetation regrowth, leading to False Positives.
- In Brisighella FAA, a series of False Positives are mapped, primarily due to the challenging task of mapping a different type of landslide and lithology (Blue Clays) not present in the training dataset.

These difficulties are relatively common across different municipalities and among various combinations of input layers used to train the models (Tab. 5.1). Regarding the differences between UNet and Segform, despite similarities in scores and mapping, Segform, being a more recent and advanced model, demonstrates a better ability to leverage input data, including those with evidently lower information content. In contrast, UNet tends to produce more homogeneous and continuous mapping, allowing it to achieve good scores.

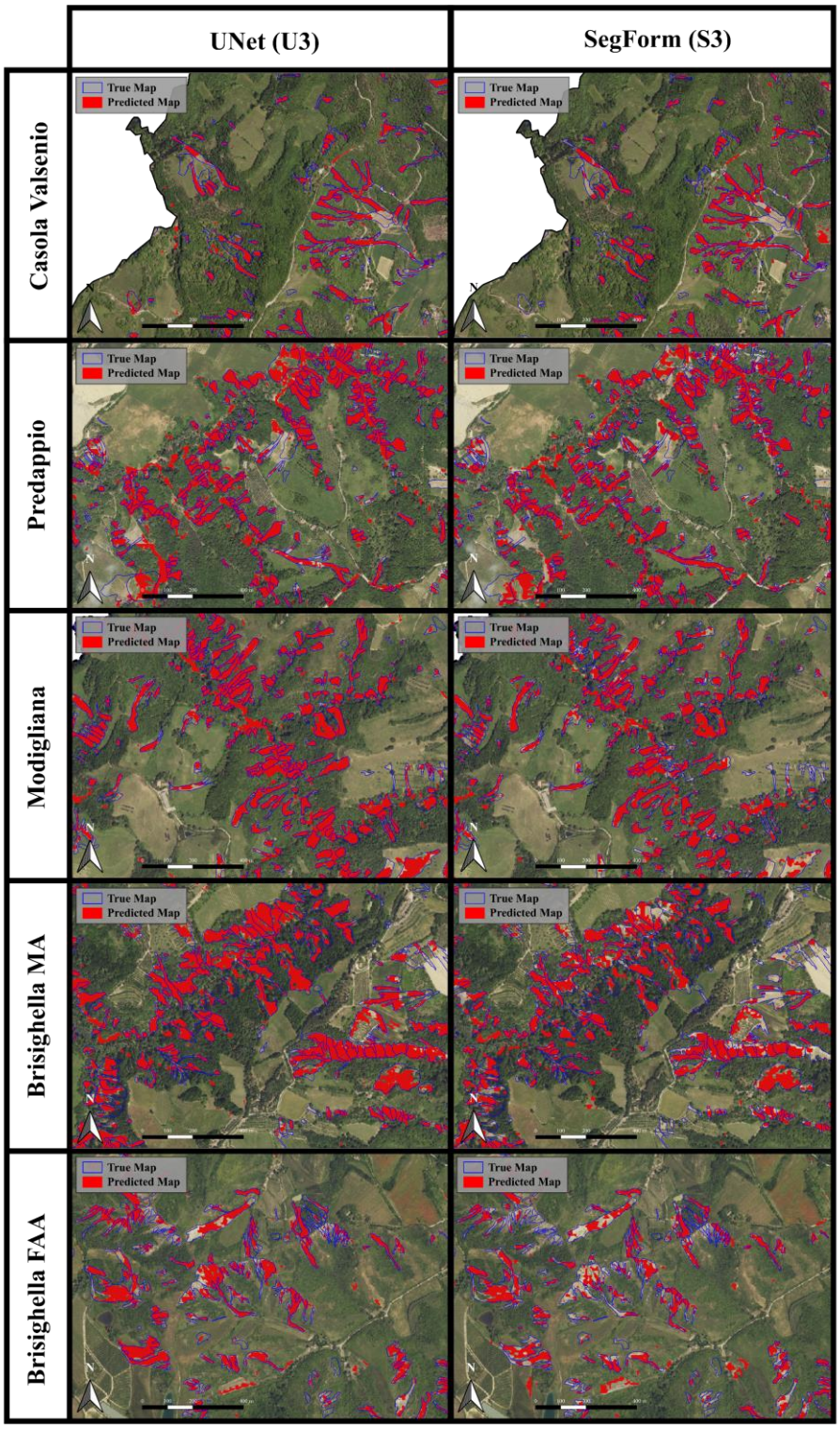


Figure 5.7. Comparative analysis of landslide mapping results using U3 and S3 models across different municipalities. The figure illustrates the spatial distribution and extent of landslides as identified by the U3 (UNet-based) and S3 (Segform-based) models in Casola Valsenio, Predappio, Modigliana, Brisighella MA, 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).

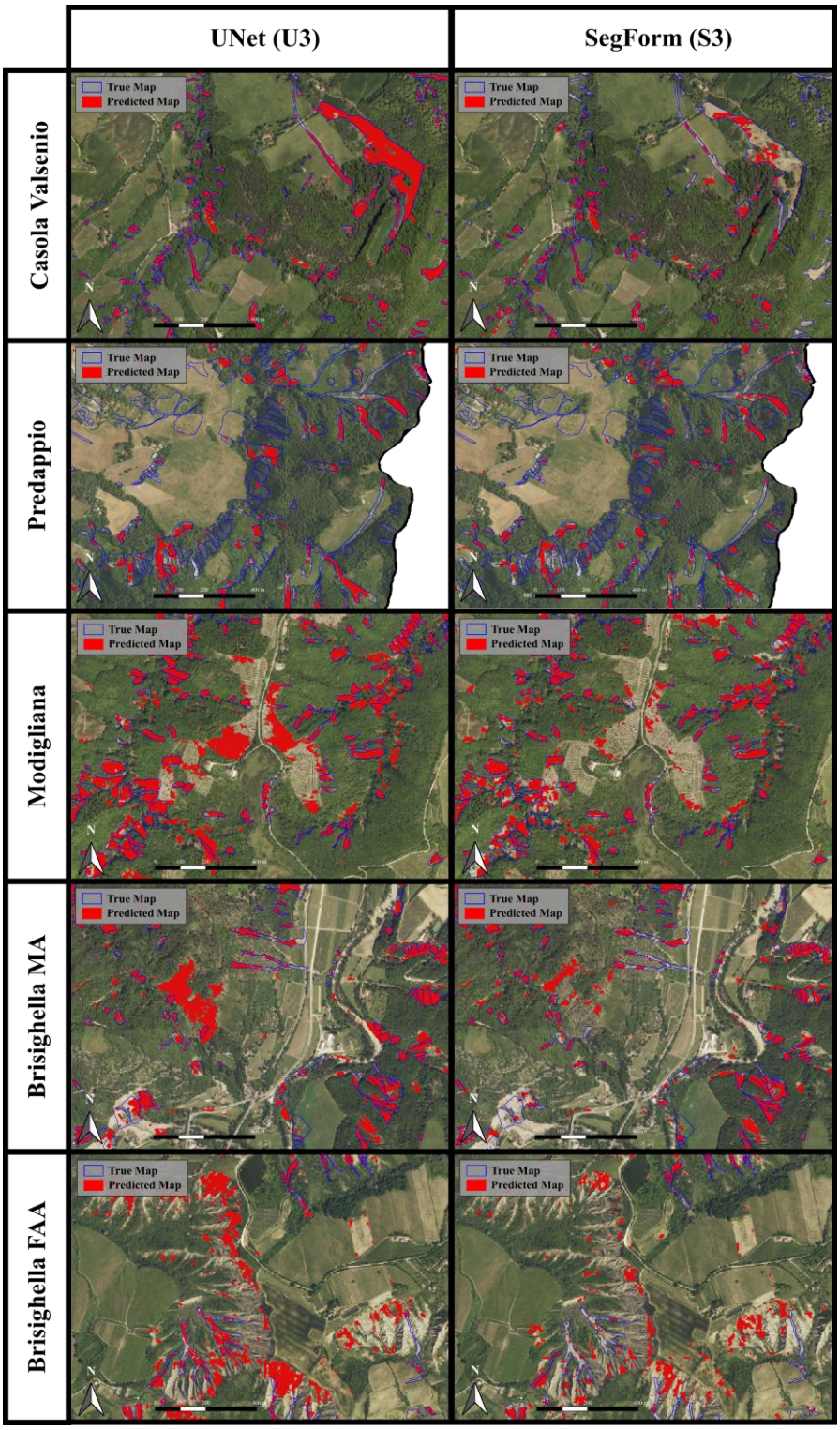


Figure 5.8. Representative examples of challenging scenarios for landslide detection using U3 and S3 models across different municipalities. This figure illustrates common situations where both models encounter difficulties in accurately mapping landslides, highlighting the limitations and potential areas for improvement in the machine learning approach. The examples are drawn from Casola Valsenio, Predappio, Modigliana, Brisighella MA, and Brisighella FAA, showcasing the diverse geological and environmental conditions that can impact model performance.

5.7 Discussion

The evaluation of automated landslide mapping techniques in this study incorporated both quantitative metrics and expert qualitative assessment, providing a comprehensive understanding of the models' performance and practical utility.

1. As outlined in section 3.7, expert evaluation played a crucial role in assessing the real-world applicability of the automated mappings. Two experienced geologists independently reviewed the output maps for the Modigliana municipality, categorizing them into three classes: low, medium, and high quality. This classification was based on several key factors critical for emergency response scenarios, including overall mapping accuracy, false positive detections in agricultural fields and along river margins, misclassifications of buildings and roads, and the shape and delineation of landslide polygons. This expert-driven approach provided valuable insights into the practical usability of automated mappings, going beyond the capabilities of quantitative metrics alone.
2. Comparing the quantitative metrics (F1 score and IoU) reported in Table 2 with the expert evaluations revealed a strong correlation between these two assessment methods. This alignment validates the reliability of our quantitative metrics and underscores their relevance in practical applications. Notably, the models U4 and S4, which were trained using all available input layers except the Sentinel-2 NDVI change data (as detailed in Table 1), received the highest expert ratings for Modigliana and also achieved the highest F1 and IoU scores. This consistency was also observed at the lower end of the performance spectrum, where models U1 and S2 (trained with RGB+Slope and RGB+Slope+pre-event imagery, respectively) were identified by experts as the poorest performers, aligning with their lower quantitative scores.
3. It is worth noting that there was one discrepancy between expert evaluation and quantitative metrics. The S5 model, trained using all available input layers (Table 1), received a low expert score despite achieving a relatively high F1 score of 0.53. This discrepancy highlights the importance of combining quantitative metrics with expert judgment, as some aspects of mapping quality that are crucial for practical applications may not be fully captured by numerical scores alone.
4. The utility of these automated mapping techniques in emergency scenarios is significant. The ability to rapidly produce landslide maps of comparable quality to expert-generated ones could substantially expedite the initial response phase. By eliminating a significant portion of the manual polygon creation work, these automated methods could save weeks of effort. This time-saving aspect is particularly crucial in emergency situations, where rapid response is essential. The automated maps could serve as a valuable starting point, allowing responders to focus their efforts on refining and verifying the results, rather than

starting from scratch. This approach could lead to the production of a final, high-quality map in a fraction of the time required for fully manual mapping.

5. It's important to note that while the F1 scores achieved by our models may not appear exceptionally high when compared to some machine learning benchmarks, they represent a significant achievement in the context of landslide mapping. Unlike many image classification tasks, landslide mapping requires precise delineation of complex, irregular shapes across varied terrain. Our models aim to produce a "real" mapping of landslide extents, including subtle transitions and complex morphologies, rather than simplified or generalized representations. This level of detail and accuracy is challenging to achieve but is crucial for practical applications in disaster response and risk assessment.
6. Finally, regarding the practical application of this methodology, the emergency commission identified buildings within the landslide boundaries and an additional 20-meter buffer zone as potentially at risk. To assess the potential utility of our automated mapping products during emergency response, we compared the buildings identified by the automated mapping with those identified through manual mapping (as shown in Figure 5.9). The F1 scores indicate that the "U4" case remains the most accurate mapping (as suggested by previous F1 scores and expert judgment), successfully identifying 227 out of 351 buildings at risk. While these results are not excellent, they demonstrate the model's capacity to accurately assess the scale of the problem. This is critical for estimating the required workforce on the ground and evaluating potential economic damages—an urgent need following the May 2023 disaster, as outlined in the European Commission's proposal for financial assistance through the European Union Solidarity Fund (European Commission, 2024).

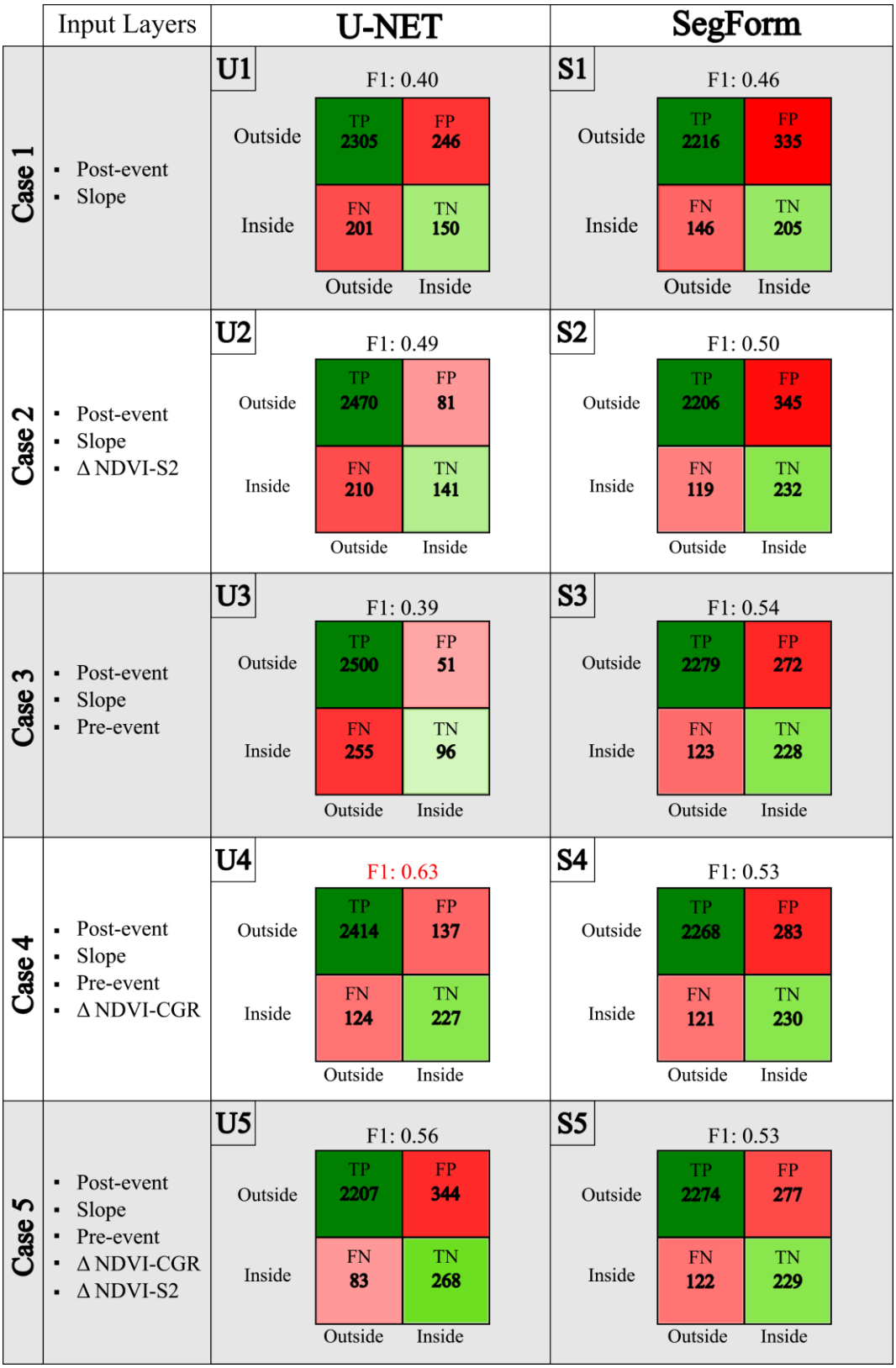


Figure 5.9. Comparison of buildings identified as at risk (within landslide boundaries or within a 20-meter buffer) by automated mapping methods and manual mapping (ground truth), showing the confusion matrices for all five cases evaluated.

5.8 Conclusion

Automated landslide mapping presents considerable advantages during emergencies. While it cannot fully replace manual mapping, it serves as a valuable tool for initial assessments. By automating the creation of landslide polygons, field operators can significantly reduce the time spent on this task and focus instead on refining and correcting model outputs. This approach streamlines the overall mapping process, which is critical during disaster response phases.

Experts generally agreed with the F1 score results, identifying consistent patterns of false positives, especially in areas like agricultural fields or along riverbanks. These errors are easily corrected during manual review, making the model outputs highly useful as a starting point. Automated mapping accelerates the creation of these foundational maps, enabling quicker responses and resource allocation in disaster situations.

In future disasters of similar scale, we have shown that machine learning models can be instrumental in expediting post-disaster assessments. Although these tools are not a substitute for thorough manual mapping, they can significantly shorten the time needed to produce accurate maps, which is vital for effective disaster management. This increased efficiency can also aid in more timely damage assessments, which are critical for securing financial support, such as the assistance provided by the European Union Solidarity Fund (European Commission, 2024).

5.9 References

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

Berti, M., Pizziolo, M., Scaroni, M., Generali, M., Olivucci, S., Gozza, G., et al. (2024a). Emergency mapping: lessons learned from 2023 landslide event in Romagna. Manuscript submitted for publication in *Landslides*.

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., and Corsini, A. (2024b). RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event, *Earth Syst. Sci. Data* Discuss. [preprint], https://doi.org/10.5194/essd-2024-407, in review.

Brath A., Casagli N., Marani M., Mercogliano P., Motta R. (2023). Rapporto della Commissione tecnico-scientifica istituita con deliberazione della Giunta Regionale n. 984/2023 e determinazione dirigenziale 14641/2023, al fine di analizzare gli eventi meteorologici estremi del mese di maggio 2023. Technical Report, Regione Emilia-Romagna, 147 pp. https://www.regione.emilia-romagna.it/alluvione/rapporto-della-commissione-tecnico-scientifica.

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.

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

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

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/.

Ferrario, M. F., & Livio, F. (2024). 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., 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

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, 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

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

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.

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

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

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

Prakash N. and Manconi A. (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).

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.

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.

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*, 1 4, 2884. https://doi.org/10.3390/rs1412f2884

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

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).

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.

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

Chapter 6

General conclusions

This PhD research aimed to explore how machine learning (ML) techniques can improve landslide prediction and emergency management, with a focus on the Emilia-Romagna region. The goal was to address three fundamental research questions, examining the application of ML methods in complex geological conditions, their effectiveness in emergency mapping, and the integration of these methods into operational early warning systems. Below, we summarize the key findings from these chapters before addressing the specific research questions.

6.1 Summary of Key Findings

Comparative Performance of Different Methods

Chapter 2 highlighted the potential of ML methods, particularly XGBoost, to enhance the prediction of rainfall-induced landslides. Compared to traditional statistical approaches like BART, XGBoost achieved notable improvements, such as reducing false positives and achieving higher overall accuracy. These results emphasize that ML methods depend heavily on the quality and specificity of input data, as the additional parameters tested had limited impact on performance. Challenges such as imprecise landslide initiation dates and varying local responses to rainfall further constrained the models' predictive capabilities.

Despite these constraints, the findings demonstrated that ML methods, when applied to well-curated datasets, offer advantages over conventional techniques by processing complex, multi-variable relationships. However, achieving meaningful advancements requires prioritizing data quality over model complexity. Refining historical datasets and improving the temporal and spatial resolution of rainfall and landslide data are critical steps to realizing the full potential of ML methods.

Machine Learning versus Manual Mapping

The analysis in Chapter 3, 4, and Chapter 5 highlighted the advantages and limitations of using ML methods compared to manual mapping in emergency scenarios. ML approaches, particularly U-Net, demonstrate significant potential in automating landslide detection and reducing the time required for initial mapping efforts. In the May 2023 Emilia-Romagna disaster, U-Net applied to high-resolution imagery achieved an F1-score of 0.64, successfully identifying over 65% of damaged roads and buildings.

However, ML methods often require manual refinement to address misclassifications, such as rooftops or agricultural fields mistakenly labeled as landslides. While high-resolution imagery helps mitigate these inaccuracies, it cannot entirely eliminate them. Sentinel-2, although

valuable for broad assessments, struggles to capture smaller landslides, highlighting the importance of data quality and resolution.

Manual mapping, though time-intensive, remains indispensable for validating ML outputs, resolving ambiguities, and refining critical details. This hybrid approach—where ML provides a baseline map and experts refine it—has proven effective in accelerating mapping efforts and improving accuracy. Moving forward, combining ML's speed with the interpretative strength of human expertise offers the most promising pathway for landslide mapping in emergency contexts.

Remaining Challenges for ML Applications in Emergency Contexts

Despite their promise, ML methods face significant challenges when applied in emergency contexts, as highlighted in Chapter 3, 4 and Chapter 5. Their performance depends heavily on the quality and availability of input data. While high-resolution imagery is critical for reliable results, its timely acquisition remains a challenge in many emergencies. Freely available Sentinel-2 imagery, with its 10-meter resolution, is insufficient for detecting small landslides or accurately delineating boundaries.

False positives and false negatives persist as key issues. Misclassifications, such as rooftops or agricultural fields erroneously identified as landslides, can overestimate damage, while undetected landslides obscured by vegetation may lead to critical oversights. These limitations emphasize the need for manual intervention to refine automated outputs, underscoring the current gaps in fully autonomous mapping.

Additionally, the "black-box" nature of some ML models, such as U-Net, limits their interpretability and acceptance in operational workflows. This lack of transparency can hinder decision-making, especially in high-pressure scenarios where justifications for predictions are required. Harmonizing classification criteria and developing user-friendly interfaces are crucial to improving both the usability and reliability of ML methods.

Finally, practical integration into emergency management systems remains challenging. Agencies often lack the computational resources and expertise needed to deploy ML models effectively. Addressing this gap through targeted training, enhanced interpretability, and streamlined workflows is essential for leveraging ML's potential in real-world applications.

6.2 Addressing the Research Questions

Question 1: *How can we effectively apply ML methods in complex geological conditions?*

The first research question sought to determine whether ML techniques could overcome the challenges posed by complex geological conditions when predicting landslide events. Chapter

2 explored this issue by applying several ML methods, to define rainfall thresholds for landslides in the Emilia-Romagna region, a setting characterized by intricate geological structures and variable topography. The results demonstrated that ML models outperformed traditional statistical methods such as Frequentist and Percentile approaches. XGBoost, for example, showed improvements in predictive accuracy and a reduction in false positives.

However, the study also revealed that these improvements were relatively modest, especially considering the additional parameters used, such as antecedent rainfall over periods of 7 to 60 days, the day of the year, and the alert areas. This suggests that while ML methods hold promise, their effectiveness is closely tied to the accuracy and resolution of the input data. The main challenge remains in accurately predicting individual rainfall events that trigger landslides, especially in regions with overlapping events where the distinction between critical and non-critical conditions is blurred.

In conclusion, ML methods can indeed improve landslide prediction in complex geological settings, but their success is limited by the availability of detailed and clean datasets, especially when dealing with historical data. The key to "improving" these predictions lies not only in using more advanced models but mainly in refining the quality of the data used to train these models. Future research should prioritize systematic monitoring and collection of high-quality, ground-truth data.

Question 2: *Is it truly effective to use ML methods for mapping during emergency situations?*

The research in Chapter 3, 4 and Chapter 5 demonstrates that machine learning (ML) models, particularly U-Net, can be highly effective for landslide mapping during emergencies, especially in large-scale disasters. One key advantage is the significant reduction in time required for mapping large numbers of landslides, compared to traditional manual methods such as delineating individual polygons in GIS. Manual mapping is labor-intensive (Chapter 4), making it not always suitable in urgent situations where a rapid response is crucial.

ML models like U-Net and SegFormer proved invaluable during large-scale events, such as the May 2023 Emilia-Romagna disaster, by enabling the rapid generation of comprehensive maps that would have taken months to create manually. While the accuracy of these models is highly dependent on input data resolution, high-resolution imagery (0.2 m) typically yields significantly superior results compared to lower resolutions (10 m, like Sentinel-2). Sentinel-2 can effectively map large-scale landslides, it presents challenges for reliable mapping in emergency scenarios where precision is essential. Thus, prioritizing the acquisition of high-resolution images during emergency responses is crucial for achieving accurate results.

Despite achieving high accuracy scores, fully automated methods cannot be relied upon for critical decisions, as they may lack the precision needed for tasks such as delineating landslide

boundaries or assessing infrastructure damage. However, they offer a reliable starting point that allows experts to focus on fine-tuning, ensuring accuracy and reliability. This hybrid approach accelerates the initial mapping while enabling quick expert adjustments, ultimately delivering credible results in much less time than manual mapping alone.

Question 3: *How can ML models be integrated into operational early warning and emergency response systems?*

Integrating machine learning (ML) models into operational early warning and emergency response systems augments existing frameworks by providing deeper insights and improving decision-making processes. Rather than replacing conventional methods, our approach synergistically combines ML with traditional tools to create a more effective and holistic system.

In the Emilia-Romagna Region, the Civil Protection Agency employs reference rainfall thresholds, numerical models for meteorological and hydrological forecasts, and real-time monitoring systems, as outlined in the Regional Council Decree No. 1761 (dated November 30, 2020). These systems use a predefined color-coded alert mechanism to assess weather and landslide risks. To complement these established methods, a platform was designed and developed by the author of this thesis, integrating the best-performing ML models into a unified tool that offers real-time rainfall data, predictions, and ML-based risk assessments. This integration provides a fuller view of potential hazards and aids in emergency management decisions.

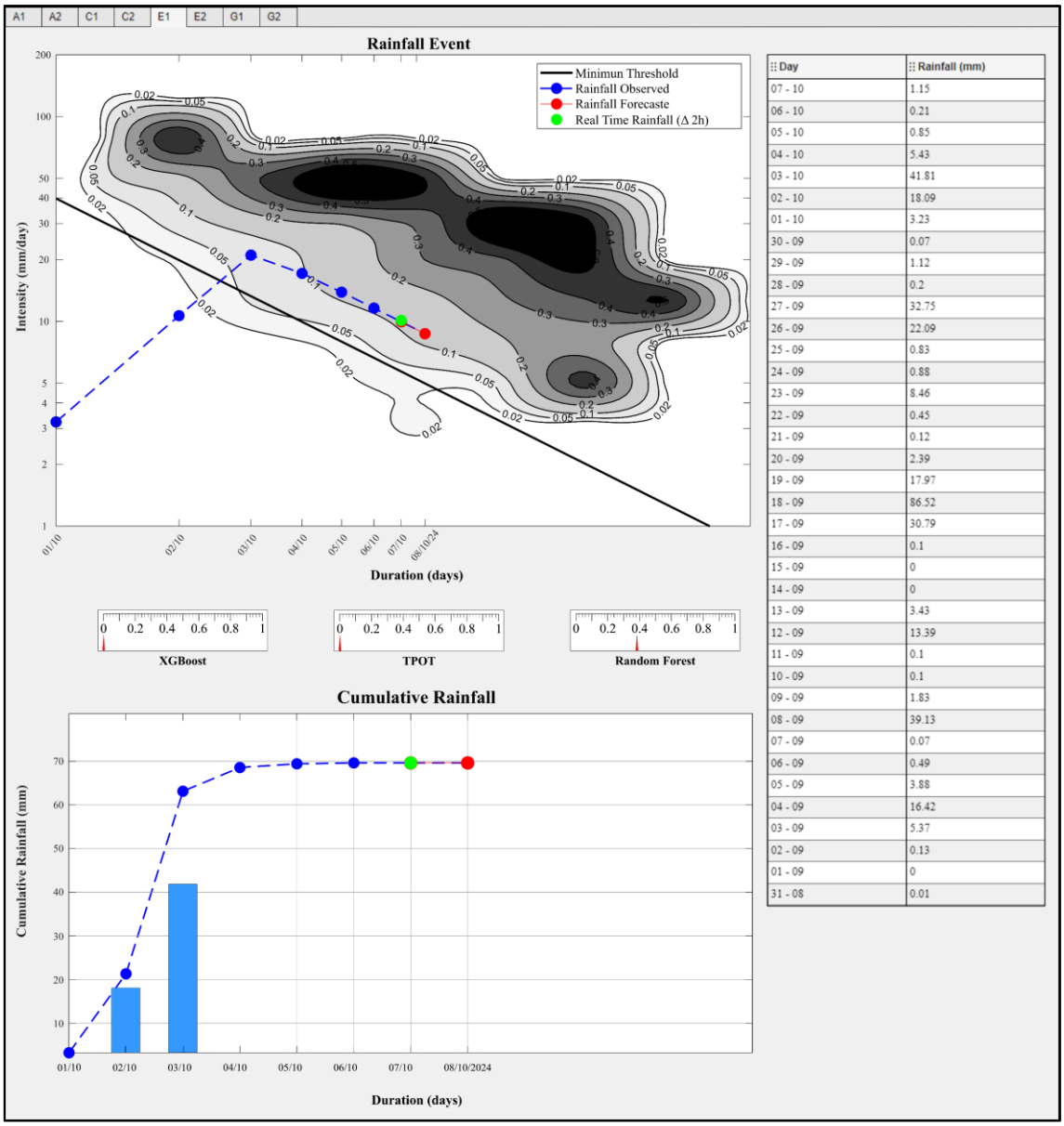
The developed graphical interface incorporates ML predictions alongside existing tools, updating in real time (Fig. 6.1, 6.2). Operators can visualize ongoing rainfall events defined by the D3P5 criteria (see section 2.5.1) within selected macroareas (see Fig. 2.1). The interface displays Bayesian probabilities of rainfall-triggered landslides, as calculated by Berti et al. (2012, 2015), along with observed rainfall data (with a two-hour latency) and forecasts for the following day (respectively available at simc.arpae and dati.arpae.). Crucially, it integrates ML predictions trained on historical data (section 2.5.1), improving the operators' ability to assess risks comprehensively. Figures 6.1 and 6.2 demonstrate the functionality of our system interface in handling with example events. Figure 6.1 displays real-time data for today's event (07/10/2024), providing operators with up-to-date information to monitor the ongoing situation within the selected macroarea. Figure 6.2, on the other hand, presents data from a significant rainfall event that occurred on 08/11/2023, which was also monitored and displayed in real time during that period. This illustrates how the system projects probabilities of triggering landslides using various algorithms.

ML models enrich the decision-making framework by adding a complementary layer of analysis to traditional methods. While operators trust established approaches like rainfall

thresholds and Bayesian analysis due to their transparency and familiarity, integrating ML provides advanced analytics and alternative perspectives. This combination enables operators to cross-check results, increasing confidence when different models concur and prompting further investigation when discrepancies arise, ultimately leading to more nuanced risk assessments.

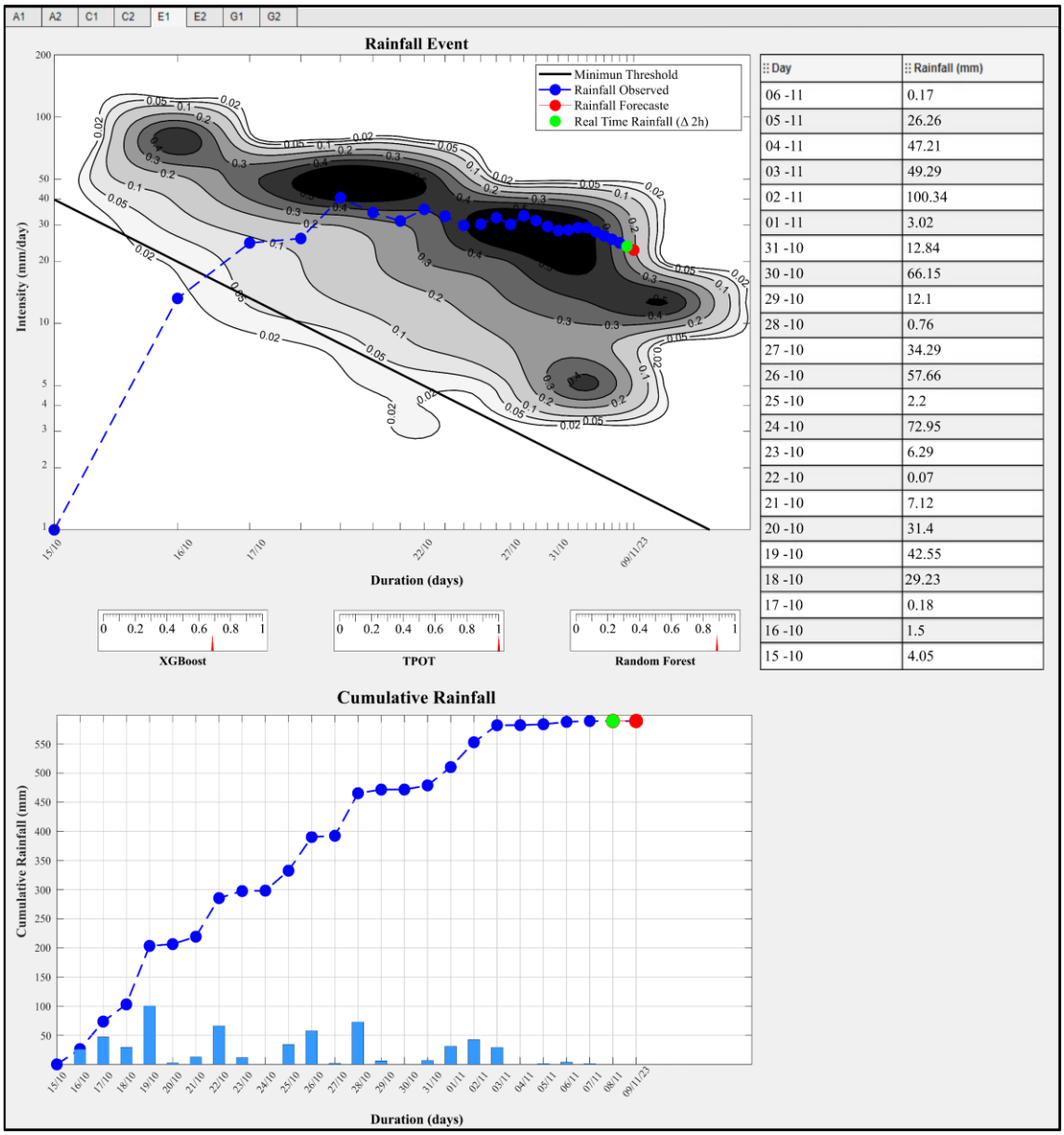
Although our platform operates independently, the Civil Protection Agency's operators could utilize this information to make informed decisions, which they then publish on the official weather alert website (allertameteo.regione.emilia-romagna) for public consultation. By enhancing internal decision-making tools, we support and align with the region's broader efforts in disaster preparedness and response.

In conclusion, integrating ML models into early warning and emergency response systems strengthens existing practices by adding analytical depth without supplanting them. This synergy between conventional techniques and ML supports more informed, timely decisions, improving the effectiveness of disaster management. By combining traditional expertise with advanced technologies, we aim to build more resilient communities capable of responding effectively to natural hazards.



Day	Rainfall (mm)
07 - 10	1.15
06 - 10	0.21
05 - 10	0.85
04 - 10	5.43
03 - 10	41.81
02 - 10	18.09
01 - 10	3.23
30 - 09	0.07
29 - 09	1.12
28 - 09	0.2
27 - 09	32.75
26 - 09	22.09
25 - 09	0.83
24 - 09	0.88
23 - 09	8.46
22 - 09	0.45
21 - 09	0.12
20 - 09	2.39
19 - 09	17.97
18 - 09	86.52
17 - 09	30.79
16 - 09	0.1
15 - 09	0
14 - 09	0
13 - 09	3.43
12 - 09	13.39
11 - 09	0.1
10 - 09	0.1
09 - 09	1.83
08 - 09	39.13
07 - 09	0.07
06 - 09	0.49
05 - 09	3.88
04 - 09	16.42
03 - 09	5.37
02 - 09	0.13
01 - 09	0
31 - 08	0.01

Figure 6.1. Illustrates the rainfall event that occurred from October 1st to October 7th, 2024, as represented in the developed graphical interface. This event is depicted using both Intensity-Duration and Cumulative-Duration graphs to provide an accurate and detailed representation of its characteristics. The graphs allow operators to assess the intensity and accumulation of rainfall over the duration of the event, which are critical factors in evaluating landslide risk. At the center of the interface, three gauges display the probabilities calculated by the machine learning algorithms XGBoost, TPOT, and Random Forest. Each gauge indicates the likelihood that this specific rainfall event will trigger a landslide in the macroarea 'E1'. The probabilities derived from these algorithms offer operators additional insights by quantifying the risk based on historical data and learned patterns. By presenting these probabilities prominently, the interface enables operators to make informed decisions quickly, enhancing the effectiveness of the early warning system.



Day	Rainfall (mm)
06 -11	0.17
05 -11	26.26
04 -11	47.21
03 -11	49.29
02 -11	100.34
01 -11	3.02
31 -10	12.84
30 -10	66.15
29 -10	12.1
28 -10	0.76
27 -10	34.29
26 -10	57.66
25 -10	2.2
24 -10	72.95
23 -10	6.29
22 -10	0.07
21 -10	7.12
20 -10	31.4
19 -10	42.55
18 -10	29.23
17 -10	0.18
16 -10	1.5
15 -10	4.05

Figure 6.2. Depicts a particularly intense rainfall event from October 15th to November 8th, 2023, as visualized in the developed graphical interface. In this case, all three models (XGBoost, TPOT, and Random Forest) indicate a probability higher than 0.7 or 70% that this event will trigger a landslide in the 'E1' macroarea.

6.2 General References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160.

Adriano, B., Yokoya, N., Miura, H., & Matsuoka, M., & Koshimura, S. (2020). A Semiautomatic Pixel-Object Method for Detecting Landslides Using Multitemporal ALOS-2 Intensity Images. *Remote Sensing*, 12(3), 561:1-19. https://doi.org/10.3390/rs12030561

Aleotti, P. (2004). A warning system for rainfall-induced shallow failures, *Eng. Geol.*, 73, 247–265. https://doi.org/10.1016/j.enggeo.2004.01.007.

Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ*., 58, 21–44. https://doi.org/10.1007/s100640050066.

Alpaydin, E. (2010). Introduction to Machine Learning, 2nd edn. MIT Press, Cambridge, MA.
Amatya, P., Kirschbaum, D., & Stanley, T. (2021). Landslide Mapping Using Object-Based Image Analysis and Open Source Tools. *Engineering Geology*, 282:1-10. https://doi.org/10.1016/j.enggeo.2021.106000.

Amatya, P., Kirschbaum, D., Stanley, T., Tanyas, H. (2021). Landslide mapping using object-based image analysis and open source tools. *Engineering Geology*, 282, 106000. https://doi.org/10.1016/j.enggeo.2021.106000

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

Angeli, M.G., Pasuto, A., & Silvano, S. (2000). A critical review of landslide monitoring experiences. *Engineering Geology*, 55(3), 133-147. https://doi.org/10.1016/S0013-7952(99)00122-2

Ardizzone, F., Bucci, F., Cardinali, M., Fiorucci, F., Pisano, L., Santangelo, M., Zumpano, V. (2023). Geomorphological landslide inventory map of the Daunia Apennines, southern Italy. *Earth Syst. Sci. Data*, 15, 753-767. https://doi.org/10.5194/essd-15-753-2023

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.

Auflič, M.J., Bezak, N., Šegina, E., Frantar, P. Gariano, S.L., Medved, A., Peternel, T. (2023). Climate change increases the number of landslides at the juncture of the Alpine, Pannonian and Mediterranean regions. *Sci Rep*, 13, 23085. https://doi.org/10.1038/s41598-023-50314-x

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.1409.0473

Ball, N.M., & Brunner, R.J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07), 1049-1106. https://doi.org/10.1142/S0218271810017160

Baum, R.L., & Godt, J.W. (2010). Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides*, 7(3), 259-272. https://doi.org/10.1007/s10346-009-0177-0

Behling, R., Roessner, S., Kaufmann, H., & Kleinschmit, B. (2014). Automated Spatiotemporal Landslide Mapping over Large Areas Using RapidEye Time Series Data. *Remote Sens* (Basel) 6:8026–8055. https://doi.org/10.3390/rs6098026

Behling, R., & Roessner, S. (2017). Spatiotemporal Landslide Mapper for Large Areas Using Optical Satellite Time Series Data. In: Advancing Culture of Living with Landslides. *Springer International Publishing*, Cham:143–152.

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Computational Biology*, 4(10), e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166. https://doi.org/10.1109/72.279181

Berti, M., Martina, M.L.V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface*, 117(F4). https://doi.org/10.1029/2012JF002367

Berti, M., Martina, M.L.V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2015). Implementation of a Probabilistic Model of Landslide Occurrence on a Civil Protection

Alert System at Regional Scale. In: Lollino, G., et al. *Engineering Geology for Society and Territory* - Volume 2. Springer, Cham. https://doi.org/10.1007/978-3-319-09057-3_110

Berti, M., Pizziolo, M., Scaroni, M., Generali, M., Olivucci, S., Gozza, G., et al. (2024a). Emergency mapping: lessons learned from the 2023 landslide event in Romagna. Manuscript submitted for publication in *Landslides*.

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., and Corsini, A. (2024b). RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna event, *Earth Syst. Sci. Data* Discuss. [preprint], https://doi.org/10.5194/essd-2024-407, in review.

Bertolini, G., Guida, M., & Pizziolo, M. (2005). Landslides in Emilia-Romagna region (Italy): strategies for hazard assessment and risk management. *Landslides*, 2, 302–312. https://doi.org/10.1007/s10346-005-0020-1

Bertolini, G., Corsini, A., Tellini, C. (2017). Fingerprints of Large-Scale Landslides in the Landscape of the Emilia Apennines. In: *Landscapes and Landforms of Italy*. World Geomorphological Landscapes, edited by Soldati, M. and Marchetti, M., 215-224, Springer, Cham. https://doi.org/10.1007/978-3-319-26194-2_18

Bhowmick, A., & Hazarika, S.M. (2018). Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends. https://doi.org/10.48550/arXiv.1606.01042

Bhuyan, K., Tanyaş, H., Nava, L. Puliero, S., Meena, S.R., Floris, M., van Westen, C., Catani, F. (2023). Generating multi-temporal landslide inventories through a general deep transfer learning strategy using HR EO data. *Sci Rep*, 13, 162. https://doi.org/10.1038/s41598-022-27352-y

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bonham-Carter G.F. (1994). Geographic Information Systems for Geoscientists, Modelling with GIS. *Pergamon Press*, Oxford. ISBN 978-0-08-041867-4.

Brath A., Casagli N., Marani M., Mercogliano P., Motta R. (2023). Rapporto della Commissione tecnico-scientifica istituita con deliberazione della Giunta Regionale n. 984/2023 e determinazione dirigenziale 14641/2023, al fine di analizzare gli eventi meteorologici estremi del mese di maggio 2023. Technical Report, Regione Emilia-Romagna, 147 pp. https://www.regione.emilia-romagna.it/alluvione/rapporto-della-commissione-tecnico-scientifica.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). Classification and regression trees. *CRC press*.

Brown, C. E. (1998). Multiple Logistic Regression. *Applied Multivariate Statistics in Geohydrology and Related Sciences,* 147–151. doi:10.1007/978-3-642-80328-4_12

Brown, T.B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Brunetti, M.T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in Italy. *Nat Hazards Earth Syst Sci* 10:447–458. https://doi.org/10.5194/nhess-10-447-2010

Brynjolfsson, E. and Mcafee, A. (2017). The Business of Artificial Intelligence. *Harvard Business Review*, 7, 3-11. https://starlab-alliance.com/The-Business-of-Artificial-Intelligence.pdf (accessed 14 Aug 2024)

Burgess, D.W., Lewis, P., & Muller, J.P. (1995). Topographic effects in AVHRR NDVI data. *Remote Sens. Environ*., 54(3), 223–232. https://doi.org/10.1016/0034- 4257(95)00155-7

Calvello, M., & Piciullo, L. (2016). Assessing the performance of regional landslide early warning models: The EDuMaP method. *Natural Hazards and Earth System Sciences*, 16(1), 103–122. https://doi.org/10.5194/nhessd-3-6021-2015

Caine, N. (1980). The rainfall intensity-duration control of shallow landslides and debris flows. *Geogr. Ann., Ser. A*, 62(1–2), 23–27. https://doi.org/10.2307/520449

Cardinali, M., Galli, M., Guzzetti, F., Ardizzone, F., Reichenbach, P., Bartoccini, P. (2006). Rainfall induced landslides in December 2004 in south-western Umbria, central Italy: types, extent, damage and risk assessment. *Natural Hazards and Earth System Sciences* 6, 237-260. https://doi.org/10.5194/nhess-6-237-2006

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.

Catani F. (2021). Landslide detection by deep learning of non-nadiral and crowdsourced optical images. *Landslides*, 18, 1025-1044. https://doi.org/10.1007/s10346-020-01513-4

Chandra, N., Sawant, S., & Vaidya, H. (2023). An Efficient U-Net Model for Improved Landslide Detection from Satellite Images. PFG - Journal of Photogrammetry, *Remote Sensing and Geoinformation Science*, 91, 13–28. https://doi.org/10.1007/s41064-023-00232-4

Chapelle, O., Scholkopf, B., & Zien, A. (2006). Semi-Supervised Learning. *MIT Press*.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785–794. https://doi.org/10.1145/2939672.2939785

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

Chen, L., et al. (2021). Decision transformer: Reinforcement learning via sequence modeling. *Advances in Neural Information Processing Systems*, 34, 15084-15097. https://doi.org/10.48550/arXiv.2106.01345

Chen, H., He, Y., Zhang, L., et al. (2023). A landslide extraction method of channel attention mechanism U-Net network based on Sentinel-2A remote sensing images. *Int J Digit Earth* 16, 552–577. https://doi.org/10.1080/17538947.2023.2177359

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).* https://doi.org/10.48550/arXiv.1406.1078

Ciccarese, G., Mulas, M., Alberoni, P.P., Truffelli, G., Corsini, A. (2020). Debris flows rainfall thresholds in the Apennines of Emilia-Romagna (Italy) derived by the analysis of recent severe rainstorms events and regional meteorological data. *Geomorphology*, 358, 107097. https://doi.org/10.1016/j.geomorph.2020.107097

Corominas, J. (2000) Landslides and Climate. *Proceedings of the 8th International Symposium on Landslides*, Cardiff, 26-30 June 2000, 1-33.

Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.; Fotopoulou, S. (2014). Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; Pitilakis, K.; Winter, M.; Pastor,

M.; Ferlisi, S.; Tofani, V.; Hervás, J.; Smith J.: Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ*, 73, 209-263. https://doi.org/10.1007/s10064-013-0538-8

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018

Corsini, A., Ciccarese, G., Diena, M., Alberoni, P.P., Amorati, R. (2017). Debris flows in Val Parma and Val Baganza (Northern Apennines) during the 13 october 2014 alluvial event in Parma province (Italy). *Italian Journal of Engineering Geology and Environment*, Special issue V National Congress of AIGA (Cagliari, Italy), 29-38. https://doi.org/10.4408/IJEGE.2017-01.S-03

Cox D. R., The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 20, Issue 2, July 1958, Pages 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Crosta, G.B., & Frattini, P. (2001). Rainfall thresholds for triggering soil slips and debris flow. In: *Proc. of the 2nd EGS Plinius Conference on Mediterranean Storms*: Publication CNR GNDCI, 2547, pp. 463–487.

Crosta, G.B., & Frattini, P. (2003). Distributed modelling of shallow landslides triggered by intense rainfall. *Natural Hazards and Earth System Sciences*, 3(1/2), 81-93. https://doi.org/10.5194/nhess-3-81-2003

Crozier, M. (2005). Multiple-occurrence regional landslide events in New Zealand: hazard management issues. *Landslides*, 2, 247-256. https://doi.org/10.1007/s10346-005-0019-7

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

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303-314. https://doi.org/10.1007/BF02551274

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

De Vita, P., Reichenbach, P., Bathurst, J. C., Borga, M., Crozier, G. M., Glade, T., Guzzetti, F., Hansen, A., & Wasowski, J. (1998). Rainfall-triggered landslides: a reference list. *Environmental Geology*, 35(2), 219-233. https://doi.org/10.1007/S002540050308

Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Ding, A., Zhang, Q., Zhou, X., & Dai, B. (2016). Automatic recognition of landslides based on CNN and texture change detection. *In Proceedings of the Chinese Association of Automation (YAC), Youth Academic Annual Conference* (pp. 444-448). IEEE.

Dou, J., Chang, K.-T., Chen, S., Yunus, A.P., Liu, J.-K., Xia, H., & Zhu, Z. (2015). Automatic Case-Based Reasoning Approach for Landslide Detection: Integration of Object-Oriented Image Analysis and a Genetic Algorithm. *Remote Sensing*, 7(4), 4318-4342. https://doi.org/10.3390/rs70404318

Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929

Draper, N.R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). *John Wiley & Sons*.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211. https://doi.org/10.1207/s15516709cog1402_1

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.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102, 85-98. https://doi.org/10.1016/j.enggeo.2008.03.022

Ferrario, M. F., & Livio, F. (2024). 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

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *In Proceedings of the 34th International Conference on Machine Learning* (pp. 1126-1135).

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. https://doi.org/10.1111/J.1469-1809.1936.TB02137.X.

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

Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Ann Stat* 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gaidzik, K., Ramírez-Herrera, M.T. (2021). The importance of input data on landslide susceptibility mapping. *Sci Rep*, 11, 19334. https://doi.org/10.1038/s41598-021-98830-y

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, 94, 268-289. https://doi.org/10.1016/j.geomorph.2006.09.023

Gariano, S.L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, 162, 227-252. https://doi.org/10.1016/j.earscirev.2016.08.011

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Daumé III, H., & Crawford, K. (2018). *Datasheets for datasets.* https://doi.org/10.48550/arXiv.1803.09010

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S., 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.

Ghorbanzadeh, O., Didehban, K., Rasouli, H., et al. (2020). An Application of Sentinel-1, Sentinel-2, and GNSS Data for Landslide Susceptibility Mapping. *ISPRS Int J Geoinf*, 9, 561. https://doi.org/10.3390/ijgi9100561

Ghorbanzadeh, O., Crivellari, A., Ghamisi, P., et al. (2021). A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan). *Sci Rep*, 11, 14629. https://doi.org/10.1038/s41598-021-94190-9

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. https://doi.org/10.1080/20964471.2022.2031544

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680. https://doi.org/10.48550/arXiv.1406.2661

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press*, Cambridge, MA. Available online: https://www.deeplearningbook.org/ (Accessed on 26 June 2007).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens*. Environ., 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Graves, A. (2012). Supervised Sequence Labelling with Recurrent Neural Networks. *Springer*. https://doi.org/10.1007/978-3-642-24797-2

Güemes-Peña, D., López-Nozal, C., Marticorena-Sánchez, R. et al. (2018). Emerging topics in mining software repositories. *Prog Artif Intell* **7**, 237–247. https://doi.org/10.1007/s13748-018-0147-7

Gunning, D., & Aha, D.W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Guzzetti, F., Cardinali, M., Reichenbach, P., Cipolla, F., Sebastiani, C., Galli, M., Salvati, P. (2004). Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. Engineering Geology 73, 229-245. https://doi.org/10.1016/j.enggeo.2004.01.006

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3), 239-267. https://doi.org/10.1007/s00703-007-0262-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2008). The rainfall intensity-duration control of shallow landslides and debris flows: an update. *Landslides*, 5(1), 3-17. https://doi.org/10.1007/s10346-007-0112-1

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112, 42-66. https://doi.org/10.1016/j.earscirev.2012.02.001

Handwerger, A.L., Fielding, E.J., Sangha, S.S., Bekaert, D.P.S. (2022). Landslide Sensitivity and Response to Precipitation Changes in Wet and Dry Climates, *Geophys. Res. Lett.*, 49, 1-12, https://doi.org/10.1029/2022GL099499

Haneberg, W.C., Cole, W.F., & Kasali, G. (2009). High-resolution lidar-based landslide hazard mapping and modeling, UCSF Parnassus Campus, San Francisco, USA. *Bulletin of Engineering Geology and the Environment*, 68, 263–276. https://doi.org/10.1007/s10064-009-0204-3

Hao, L., Rajaneesh, A., van Westen, C., Sajinkumar, K.S., Martha, T.R., Jaiswal, P., McAdoo, B.G. (2020). Constructing a complete landslide inventory dataset for the 2018 monsoon disaster in Kerala, India, for land use change analysis. *Earth Syst. Sci. Data*, 12, 2899-2918, https://doi.org/10.5194/essd-12-2899-2020

Hartigan, J.A., & Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. *Series C (Applied Statistics)*, 28(1), 100-108. https://doi.org/10.2307/2346830

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Hinton, G.E., Osindero, S., & Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527

Ho, T.K. (1995). Random Decision Forests. *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 278-282. https://doi.org/10.1109/ICDAR.1995.598994

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

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

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257. https://doi.org/10.1016/0893-6080(91)90009-T

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). Applied Logistic Regression (3rd ed.). *Wiley*.

Hu, X., Wu, S., Zhang, G., Zheng, W., Liu, C., He, C., ... & Zhang, H. (2021). Landslide displacement prediction using kinematics-based random forests method: A case study in Jinping Reservoir Area, China. *Engineering Geology*, 283, 105975. https://doi.org/10.1016/j.enggeo.2020.105975

Huang, L., & Xiang, L.Y. (2018). Method for Meteorological Early Warning of Precipitation-Induced Landslides Based on Deep Neural Network. *Neural Process Lett*, 48, 1243-1260. https://doi.org/10.1007/s11063-017-9778-0

Hungr, O., Leroueil, S., Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, 11, 167–194. https://doi.org/10.1007/s10346-013-0436-y

Iverson, R.M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, 36(7), 1897-1910. https://doi.org/10.1029/2000WR900090

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

Jaboyedoff, M., Oppikofer, T., Abellan, A., Pedrazzini, A. (2012). Use of LIDAR in landslide investigations: A review. *Natural Hazards*, 61, 5-28. https://doi.org/10.1007/s11069-010-9634-2

Jain, A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain, N., Roy, P., Martha, T.R., et al. (2024). Causal analysis of unprecedented landslides during July 2021 in the Western Ghats of Maharashtra, India. *Landslides*, 21, 99-109. https://doi.org/10.1007/s10346-023-02165-w

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

Jordan, M.I., & Bishop, C.M. (1996). Neural networks. In A. Tucker (Ed.), *CRC Handbook of Computer Science CRC Press.*

Jordan, M.I., & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Joubert-Boitat, I., Wania, A., & Dalmasso, S. (2020). Manual for CEMS-Rapid Mapping Products. Ispra.

Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Karthikeyan, A., Priyakumar, U.D. Artificial intelligence: machine learning for chemical sciences. *J Chem Sci* 134, 2 (2022). https://doi.org/10.1007/s12039-021-01995-2

Keyport, R.N., Oommen, T., Martha, T.R., Sajinkumar, K.S., & Gierke, J.S. (2018). A comparative analysis of pixel- and object-based detection of landslides from very high-resolution images. *Int. J. Appl. Earth Obs. Geoinf.*, 64, 1-11. https://doi.org/10.1016/j.jag.2017.08.015

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

Kirschbaum, D., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, 6(3),505-523. https://doi.org/10.1002/2017EF000715

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence* - Volume 2. Morgan Kaufmann Publishers Inc., San Francisco, pp 1137–1143. https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf (accessed 12 Dec 2023).

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. https://doi.org/10.1145/3065386

Kutner, M.H., Nachtsheim, C.J., & Neter, J. (2004). Applied Linear Regression Models. *McGraw-Hill Irwin*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. https://doi.org/10.1109/5.726791.

LeCun, Y., Bengio, Y., & Hinton, G.E. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Li, Z., Shi, W., Myint, S.W., Lu, P., & Wang, Q. (2016). Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sensing of Environment*, 175, 215-230. https://doi.org/10.1016/j.rse.2016.01.003

Liu, Z., Gilbert, G., Cepeda, J.M., Lysdahl, A.O.K., Piciullo, L., Hefre, H., & Lacasse, S. (2021). Modelling of shallow landslides with machine learning algorithms. *Geoscience Frontiers*, 12(1), 385-393. https://doi.org/10.1016/j.gsf.2020.04.014

Lloyd, S.P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137. https://doi.org/10.1109/TIT.1982.1056489

Loh, W.Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery*, 1(1), 14-23. https://doi.org/10.1002/widm.8

Luetzenburg, G., Svennevig, K., Bjørk, A.A., Keiding, M., Kroon, A. (2022). A national landslide inventory for Denmark. *Earth Syst. Sci. Data*, 14, 3157-3165. https://doi.org/10.5194/essd-14-3157-2022

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765- 4774.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281-297.

Magrì, S., Solimano, M., Delogu, F., et al. (2024). Modelling rainfall-induced landslides at a regional scale, a machine learning based approach. *Landslides*, 21, 573-582. https://doi.org/10.1007/s10346-023-02173-w

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surf. Processes*, 29, 687-711. https://doi.org/10.1002/esp.1064

Martinez, S.N., Schaefer, L.N., Allstadt, K.E., & Thompson, E.M. (2021). Evaluation of Remote Mapping Techniques for Earthquake-Triggered Landslide Inventories in an Urban Subarctic Environment: A Case Study of the 2018 Anchorage, Alaska Earthquake. *Front Earth Sci* (Lausanne), 9. https://doi.org/10.3389/feart.2021.673137

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

Menard, S. (2002). Applied Logistic Regression Analysis (2nd ed.). *Sage*.

Mikhaylov, S.J., Esteve, M., & Campion, A. (2018). Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration. *Phil. Trans. R. Soc. A*, 376:20170357. https://doi.org/10.1098/rsta.2017.0357

Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech*, 2, 1045-1048. https://doi.org/10.21437/Interspeech.2010-343

Mitchell, T.M. (1997). Machine Learning. McGraw-Hill.

Mondini, A.C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, 14(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

Mondini, A.C., Guzzetti, F., Reichenbach, P., et al. (2011). Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens Environ*, 115, 1743-1757. https://doi.org/10.1016/j.rse.2011.03.006

Moore, D.S., & McCabe, G.P. (1993). Introduction to the Practice of Statistics. *WH Freeman/Times Books/Henry Holt & Co., New York.*

Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective. *MIT Press.*

Nair, V., & Hinton, G.E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning* (ICML-10), 807-814.

Nava, L., Bhuyan, K., Meena, S.R., et al. (2022). Rapid Mapping of Landslides on SAR Data by Attention U-Net. *Remote Sens* (Basel), 14. https://doi.org/10.3390/rs14061449

Neapolitan, R.E. (2003). Learning Bayesian Networks. Upper Saddle River, NJ: Prentice Hall.

Niyokwiringirwa, P., Lombardo, L., Dewitte, O., et al. (2024). Event-based rainfall-induced landslide inventories and rainfall thresholds for Malawi. *Landslides*. https://doi.org/10.1007/s10346-023-02203-7

Nocentini, N., Rosi, A., Piciullo, L. et al. Regional-scale spatiotemporal landslide probability assessment through machine learning and potential applications for operational warning systems: a case study in Kvam (Norway). *Landslides* 21, 2369–2387 (2024). https://doi.org/10.1007/s10346-024-02287-9

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, 2625-2648. https://doi.org/10.5194/nhess-23-2625-2023

Notti, D., Cignetti, M., Godone, D., Cardone, D., Giordan, D. (2024). The unsuPervised shAllow laNdslide rapiD mApping: PANDA method applied to severe rainfalls in northeastern appenine (Italy). *International Journal of Applied Earth Observation and Geoinformation*, 129, 103806. https://doi.org/10.1016/j.jag.2024.103806

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

Olson, R.S., Bartley, N., Urbanowicz, R.J., & Moore, J.H. (2016). Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science. GECCO 2016 - *Proceedings of the 2016 Genetic and Evolutionary Computation Conference*, 485-492. https://doi.org/10.1145/2908812.2908918

Olson, R.S., Moore, J.H. (2019). TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In: Hutter, F., Kotthoff, L., Vanschoren, J. (eds) Automated Machine Learning. *The Springer Series on Challenges in Machine Learning*. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_8

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1717-1724. https://doi.org/10.1109/CVPR.2014.222

Orland, E., Roering, J.J., Thomas, M.A., & Mirus, B.B. (2020). Deep learning as a tool to forecast hydrologic response for landslide-prone hillslopes. *Geophysical Research Letters*, 47. https://doi.org/10.1029/2020GL088731

Osanai, N., Shimizu, T., Kuramoto, K. et al. (2010). Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network. *Landslides* 7, 325– 338 (2010). https://doi.org/10.1007/s10346-010-0229-5

Pan, S.J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

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).

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. https://doi.org/10.48550/arXiv.1201.0490

Petley, D.N., Bulmer, M.H., & Murphy, W. (2005). Patterns of movement in rotational and translational landslides. *Geology*, 33(2), 161-164. https://doi.org/10.1130/0091-7613.

Piacentini, D, Troiani, F, Daniele, G, Pizziolo, M. (2018). Historical geospatial database for landslide analysis: the Catalogue of Landslide OCcurrences in the Emilia-Romagna Region (CLOCkER). *Landslides*, 15, 811–822. http://dx.doi.org/10.1007/s10346-018-0962-8

Piciullo, L., Calvello, M., & Cepeda, J.M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, 179, 228-247. https://doi.org/10.1016/j.earscirev.2018.02.013

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.

Prakash, N., Manconi, A., & Loew, S. (2021). A new strategy to map landslides with a generalized convolutional neural network. *Scientific Reports*, 11(1), 9722. 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.

Pu, M., Zhao, Y., Ni, Z., et al. (2022). Spatial-Temporal Evolution and Driving Forces of NDVI in China's Giant Panda National Park. *Int J Environ Res Public Health*, 19. https://doi.org/10.3390/ijerph19116722

Qi, W., Wei, M., Yang, W., et al. (2020). Automatic mapping of landslides by the ResU-Net. *Remote Sens* (Basel), 12. https://doi.org/10.3390/rs12152487

Qu, F., Qiu, H., Sun, H., & Tang, M. (2021). Post-failure landslide change detection and analysis using optical satellite Sentinel-2 images. *Landslides*, 18, 447-455. https://doi.org/10.1007/s10346-020-01498-0

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106. https://doi.org/10.1007/BF00116251

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434. https://doi.org/10.48550/arXiv.1511.06434

Radford, A., et al. (2018). Improving language understanding by generative pre-training. OpenAI Blog.

Ramos-Cañón, A.M., Prada-Sarmiento, L.F., Trujillo-Vela, M.G., et al. (2016). Linear discriminant analysis to describe the relationship between rainfall and landslides in Bogotá, Colombia. *Landslides*, 13, 671-681.

https://doi.org/10.1007/s10346-015-0593-2

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). https://doi.org/10.1109/CVPR.2016.91

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).

Ricci Lucchi, F., & Valmori, E. (1980). Basin-wide turbidites in Miocene, over-supplied deep-sea plain: a geometrical analysis. *Sedimentology*, 27, 241-270.

Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

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.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386-408. https://doi.org/10.1037/h0042519

Rosi, A., Tofani, V., Tanteri, L., Tacconi Stefanelli, C., Agostini, A, Catani, F., Casagli, N. (2018). The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. *Landslides*, 15, 5-19. https://doi.org/10.1007/s10346-017-0861-4

Rossi, F., Witt, A., Guzzetti, F., Malamud, B.D., & Peruccacci, S. (2010). Analysis of historical landslide time series in the Emilia-Romagna Region, Northern Italy. *Earth Surface Processes and Landforms*, 35, 1123-1137.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. https://doi.org/10.1038/323533a0

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

Samek, W., Wiegand, T., & Müller, K.R. (2023). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Model. *Springer Nature*. https://doi.org/10.48550/arXiv.1708.08296

Samuel, A.L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229. https://doi.org/10.1147/rd.33.0210

Santangelo, M., Althuwaynee, O., Alvioli, M. et al. (2023). Inventory of landslides triggered by an extreme rainfall event in Marche-Umbria, Italy, on 15 September 2022. *Sci Data*, 10, 1–11. https://doi.org/10.1038/s41597-023-02336-3

Satriano, V., Ciancia, E., Filizzola, C., Genzano, N., Lacava, T., & Tramutoli, V. (2023). Landslides Detection and Mapping with an Advanced Multi-Temporal Satellite Optical Technique. *Remote Sens.*, 15, 683. https://doi.org/10.3390/rs15030683

Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

Schölkopf, B., & Smola, A.J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. *MIT press*. https://doi.org/10.7551/mitpress/4175.001.0001

Scorpio, V., Crema, S., Marra, F., Righini, M., Ciccarese, G., Borga, M., Cavalli, M., Corsini, A., Marchi, L., Surian, N., Comiti, F. (2018). Basin-scale analysis of the geomorphic effectiveness of flash floods: a study in the northern Apennines (Italy). *Sci Total Environ*, 640–641, 337–351. https://doi.org/10.1016/j.scitotenv.2018.05.252

Seber, G.A.F., & Lee, A.J. (2012). Linear Regression Analysis. *John Wiley & Sons.*

Segoni, S., Piciullo, L., & Gariano, S.L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15(8), 1483-1501. https://doi.org/10.1007/s10346-018-0966-4

Sidle, R.C., & Ochiai, H. (2006). Landslides: processes, prediction, and land use. *American Geophysical Union*. https://doi.org/10.1029/WM018

Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484-489. https://doi.org/10.1038/nature16961

Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis.

https://doi.org/10.1201/9781315140919

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. https://doi.org/10.48550/arXiv.1312.6034

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *In Advances in Neural Information Processing Systems* (pp. 4077-4087). https://doi.org/10.48550/arXiv.1703.05175

Soeters, R., Van Westen, C.J. (1996). Slope instability recognition, analysis and zonation. Landslide Types and Processes, Transportation Research Board, *U.S. National Academy of Sciences*, Special Report, 247, 129-177

Stanley T.A., Kirschbaum D.B., Benz G., Emberson R.A., Amatya P.M., Medwedeff W. and Clark M.K. (2021). Data-Driven Landslide Nowcasting at the Global Scale. *Front. Earth Sci*. 9:640043. https://doi.org/10.3389/feart.2021.673644

Steger, S., Brenning, A., Bell, R. et al. (2017). The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides*, 14, 1767-1781. https://doi.org/10.1007/s10346-017-0820-0

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. https://doi.org/10.48550/arXiv.1906.02243

Su, Z., Chow, J.K., Tan, P.S., et al. (2021). Deep convolutional neural network–based pixel-wise landslide inventory mapping. *Landslides*, 18, 1421-1443. https://doi.org/10.1007/s10346-020-01557-6

Sun, T.Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Sutskever, I., Vinyals, O., & Le, Q.V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 3104-3112.

Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction (2nd ed.). *MIT Press.*

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

Tanoli, J.I., Ningsheng, C., Regmi, A.D., & Jun, L. (2017). Spatial distribution analysis and susceptibility mapping of landslides triggered before and after Mw7.8 Gorkha

earthquake along Upper Bhote Koshi, Nepal. *Arabian Journal of Geosciences*, 10, 277. https://doi.org/10.1007/s12517-017-3026-9

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, 17, 168–192. https://doi.org/10.1016/j.aci.2018.08.003

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

Tucker, C.J. (1979). Red and Photographic Infrared linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8(2), 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Van Den Eeckhaut, M., Hervás, J., (2012). State of the art of national landslide databases in Europe and their potential for assessing susceptibility, hazard and risk. *Geomorphology* 139-140, 545–558. https://doi.org/10.1016/j.geomorph.2011.12.006

Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008. https://doi.org/10.48550/arXiv.1706.03762

Venables, W.N., & Ripley, B.D. (2002). Modern Applied Statistics with S. https://doi.org/10.1007/978-0-387-21706-2

Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. *In Advances in Neural Information Processing Systems* (pp. 3630-3638).

Wang, G., & Sassa, K. (2003). Pore-pressure generation and movement of rainfall-induced landslides: effects of grain size and fine-particle content. *Engineering Geology*, 69(1-2), 109-125. https://doi.org/10.1016/S0013-7952(02)00268-5

Werbos, P.J. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. *Harvard University*.

Wieczorek, G.F., & Glade, T. (2007). Climatic factors influencing occurrence of debris flows. Debris-Flow Hazards and Related Phenomena, 325–362. https://doi.org/10.1007/3-540-27129-5_14

Williams, J.G., Rosser, N.J., Kincey, M.E., et al. (2018). Satellite-based emergency mapping using optical imagery: experience and reflections from the 2015 Nepal earthquakes. *Natural Hazards and Earth System Sciences*, 18, 185- 205. https://doi.org/10.5194/nhess-18-185-2018

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

Wu, X., Ren, F., & Niu, R. (2014). Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ Earth Sci*, 71, 4725-4738. https://doi.org/10.1007/s12665-013-2863-4

Xi, L.; Yu, J.; Ge, D.; Pang, Y.; Zhou, P.; Hou, C.; Li, Y.; Chen, Y.; Dong, Y. (2024). SAM-CFFNet: SAM-Based Cross-Feature Fusion Network for Intelligent Identification of Landslides. *Remote Sens*, 16, 2334. https://doi.org/10.3390/rs16132334

Xian, Y., Lampert, C.H., Schiele, B., & Akata, Z. (2018). Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9), 2251-2265. https://doi.org/10.48550/arXiv.1707.00600

Xie, E., et al. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34, 12077-12090. https://doi.org/10.48550/arXiv.2105.15203

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.

Yang, B., Yin, K., Lacasse, S., & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides*, 16, 677-694. https://doi.org/10.1007/s10346-018-01127-x

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

Yang, X., Chen, L. (2010). Using multi-temporal remote sensor imagery to detect earthquake-triggered landslides. *International Journal of Applied Earth Observation and Geoinformation*, 12, 487-495. https://doi.org/10.1016/j.jag.2010.05.006

Yang, X., Zuo, X., Xie, W., Li, Y., Guo, S., & Zhang, H. (2022). A Correction Method of NDVI Topographic Shadow Effect for Rugged Terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15. https://doi.org/10.1109/JSTARS.2022.3193419

Ye, C.M., et al. (2019). Landslide detection of hyperspectral remote sensing data based on deep learning with constraints. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12, 5047-5060. https://doi.org/10.1109/JSTARS.2019.2951725

Yu, B., Chen, F., & Xu, C. (2020). Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Computers & Geosciences*, 135, 104388. https://doi.org/10.1016/j.cageo.2019.104388.

Yu, M., Yang, C., & Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences*, 8(5), 165. https://doi.org/10.3390/geosciences8050165

Zêzere, J. L. (2002). Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal), *Nat. Hazards Earth Syst. Sci*., 2, 73–82, https://doi.org/10.5194/nhess-2-73-2002

Zhao, Y., Meng, X., Qi, T., Li, Y., Chen, G., Yue, D., & Qing, F. (2022). AI-based rainfall prediction model for debris flows. *Engineering Geology*, 296, 106456. https://doi.org/10.1016/j.enggeo.2021.106456

Zhong, C., Liu, Y., Gao, P., Chen, W., Li, H., Hou, Y., & Ma, H. (2019). Landslide mapping with remote sensing: challenges and opportunities. *International Journal of Remote Sensing*, 41(4), 1555-1581. https://doi.org/10.1080/01431161.2019.1672904

Zhou, C., Yin, K., Cao, Y., et al. (2022). Development of landslide inventories and analysis of influencing factors of landslides triggered by the 2008 Wenchuan earthquake. *Landslides*, 19, 587-600. https://doi.org/10.1007/s10346-021-01814-8

Zhu, J., Zeng, Q., Zhao, Y., et al. (2023). Detecting landslide dynamics using an enhanced spatiotemporal deep learning framework. *Remote Sens* (Basel), 15. https://doi.org/10.3390/rs15040955

Zhu, Y., Groves, K., De Lange, R., & Napieralski, J. (2022). Toward Large-Scale, High-Resolution Landslide Mapping Using Convolutional Neural Networks. *Remote Sens*., 14, 3193. https://doi.org/10.3390/rs14133193

Zhu, L., Long, M., Wang, J., & Jordan, M.I. (2019). Deep transfer learning: A survey. *arXiv preprint* arXiv:1411.1792. https://doi.org/10.1109/TPAMI.2022.3158377

Zhu, X., & Goldberg, A. B. (2009). Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 3(1), 1-130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006

Zieher, T., Perzl, F., Rössel, M., Rutzinger, M., Meißl, G., Markart, G., Geitner, C. (2016). A multi-annual landslide inventory for the assessment of shallow landslide susceptibility – Two test cases in Vorarlberg, Austria, *Geomorphology*, 259, 40-54, https://doi.org/10.1016/j.geomorph.2016.02.008

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *Lecture Notes in Computer Science*, *8689*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

Acknowledgements

At the end of this journey, I would like to take a moment to thank all the people who made it possible.

First and foremost, I would like to thank my supervisor, Matteo. Who has taught me so much, not only scientific knowledge, but also valuable human lessons. The long and stimulating conversations about scientific projects, and life in general, have truly made a difference.

I would also like to thank my friends and research colleagues Alessandro, Tommaso, Giuseppe, Rodolfo, Elena, and Matteo for the wonderful moments we shared both inside and outside the university.

Additionally, I would like to express my gratitude to Sensei Wang, my supervisor during my research period in Kyoto, for their guidance and hospitality.

Outside the academic world, I am thankful to my partner, Elisa, who supports and endures all my decisions, and for being by my side throughout this PhD journey, and hopefully beyond.

I also want to thank my friends, Enrico and Francesco, with whom I've shared many great moments, especially after work.

I am grateful to my parents, Corrado, Itana, and Alessandro, for their lifelong guidance and support, and for always believing in me.

Last but not least, I want to thank Manzitek for making every return to Sestri Levante feel like an overseas vacation.