PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Escuela de Posgrado

Evaluación de modelos de segmentación semántica para el monitoreo de deslizamiento de tierra utilizando imágenes satelitales

Trabajo de investigación para obtener el grado académico de Maestro en Informática con mención en Ciencias de la Computación que presenta:

Roy Marco Yali Samaniego

Asesor:

Pablo Alejandro Fonseca Arroyo

Lima, 2024

Informe de Similitud

Yo, **Pablo Alejandro FONSECA ARROYO***,* docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada "Evaluación de modelos de segmentación semántica para el monitoreo de deslizamientos de tierra utilizando imágenes satelitales", del autor **Roy Marco YALI SAMANIEGO**, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de **13%**. Así lo consigna el reporte de similitud emitido por el software Turnitin el **22/01/2024**.

- He revisado con detalle dicho reporte y la tesis y no se advierte indicios de plagio.

- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

Lima, 22 de enero del 2024.

Apellidos y nombres del asesor:	
FONSECA ARROYO, Pablo Alejandro	
DNI: 44695174	Firma
ORCID: 0000-0002-0208-2842	

RESUMEN

En el ámbito del aprendizaje automático, un desafío persistente es la disponibilidad de datos suficientes, especialmente en tareas de visión por computadora. Este desafío se amplifica cuando se trabaja con sensores remotos, donde las bases de datos etiquetadas para abordar problemas son escasas. Este manuscrito examina críticamente el monitoreo de deslizamientos de tierra en el paisaje peruano y presenta tres contribuciones en esta dirección.

La primera contribución expande un conjunto de datos de imágenes satelitales sobre deslizamientos de tierra (Landslide4Sense) proveniente de territorios asiáticos, con 3799 imágenes debidamente etiquetadas. Reconociendo la dinámica geoespacial de Perú, se incrementó este conjunto de datos con 838 escenarios locales. Estas adiciones mantienen congruencia con el conjunto de datos original en términos de atributos y configuración, asegurando replicabilidad y escalabilidad para futuras investigaciones.

La segunda evalúa varios modelos de segmentación semántica basados en la arquitectura U-net, reforzada por la función de pérdida de Entropía Cruzada Ponderada + Dice Loss, óptima en tareas de segmentación con conjuntos de datos desequilibrados. Los resultados permiten alcanzar un F1-Score del 75.5% con la arquitectura U-net (vanilla) superando el benchmark de referencia del 71.65%.

La última contribución muestra un desarrollado integral para la adquisición de datos, procesamiento y entrenamiento/evaluación de modelos. Dado que este marco tiene el potencial de impulsar una aplicabilidad general de sistemas de segmentación a sistemas de monitoreo de deslizamientos de tierra, y de tener un alcance más amplio a la comunidad académica y partes interesadas gubernamentales en Latinoamérica y en todo el mundo.

ÍNDICE

1 INTRODUCTION

Landslides represent a salient geological hazard, posing substantive threats to human populations, natural resources, infrastructure, and property within mountainous regions[1]. These phenomena can be precipitated by various triggers, including rainfall, thawing, earthquakes, or volcanic eruptions. In Peru, landslides are notably frequent during the rainy season, often inflicting severe damage upon communities inhabiting the mountainous regions[2]. This vulnerability is further compounded by the precarious nature of local housing structures and their frequent siting within high-risk zones[3].

The occurrence of landslides is controlled by various spatial and climatic factors, making their susceptibility assessment challenging due to the diversity and amount of spatial data that must be considered. Assessment methods can be classified into qualitative (knowledge-based) and quantitative (data-based and physically based) methods[4]. Data-driven methods have historically played a valuable role in environmental remote sensing research. This approach involves inherent challenges in interpreting large volumes of data and presents significant computational challenges in both the storage and detailed analysis of geospatial information. With an increasing amount of "big data" from earth observation and rapid advances in machine learning(ML) and deep learning(DL) increase opportunities for novel methods to aid in earth environmental monitoring[5]. Landslide research based on satellite imagery is critical to understanding their mechanics and impending failure where lives, communities, or infrastructure may be at risk. These analyses, within a computer vision context, involve feature extraction, classification, segmentation and detection to identify the area, and spectral behaviour of landslides involving a change in the ground surface.

The primary objective of the present research is to develop a framework for automatic landslide segmentation and detection using a combination of Sentinel2 and Alos Palsar satellite imagery, starting from the data acquisition from the Landslide4Sense challenge[6] and how we increase the dataset size based in Sentinel-2 and Alos Palsar images with local features and showing a workflow to prepare a dataset for classification and semantic segmentation tasks in remote sensing context, then how we use different DL models to evaluate the best combination to find landslide patterns from bands combination of images, and finally, how we storage results in a cloud service. This approach seeks to achieve a pixel level classification of landslide areas in high jungle regions in Peru, one of the countries most affected by this phenomenon in the Americas[7].

This document is structured as follows. In the "Motivation" section, we examine the encouragement for the research and the significance of identifying and mapping landslide areas. "Background" provides a state-of-the-art review of various machine learning techniques applied using freely accessible satellite imagery. "Data and resources" reveals the data sources utilized in the research, along with the process of generating proprietary data using labelling tools and techniques for segmentation models. The "Methodology" section presents the different experiments conducted, loss functions, performance measurement metrics, and experiments and a description and analysis of the results, alongside a discussion comparing them with the scientific literature. "Cloud-based monitoring system framework" shows a pipeline of how we work with image time-series and how it is integrated with a cloud-based app. Finally, the "Conclusion" section presents the final remarks and conclusions drawn from the research.

2 MOTIVATION

In recent years, Earth Observation (EO) has emerged as a pivotal discipline impacting various interdisciplinary research areas such as climate change, risk and disaster management, and foresight analysis[8]. This evolution has enabled a shift from localized, small-scale analyses to global-scale investigations, powered by the computational capabilities of major cloud-based platforms[9]. This transition has facilitated the development of sophisticated analytical frameworks, capable of addressing complex environmental challenges, thus requiring advanced domain expertise.

The inception of Google Earth Engine (GEE) in 2015 marked a significant milestone in EO, providing researchers with unparalleled access to extensive spatial data repositories and robust computational resources for real-time processing[10]. Advancements in the platform have incorporated machine learning techniques like Random Forest and Support Vector Machines[11], which can be executed from JavaScript and Python. Additionally, the rise of cloud-based services and deep learning frameworks such as TensorFlow and PyTorch has enabled seamless neural network integration within GEE, creating new avenues for large-scale evaluation of natural phenomena[12].

Geographically, this study aims to identify pre and post-landslide conditions in vulnerable zones, providing valuable data for disaster prevention, resource allocation, and community interventions. From a computer vision perspective, this work highlights the development of a convolutional neural network-based model for semantic segmentation using satellite images as inputs. This model, trained on global databases and augmented with region-specific data, offers a novel approach to monitoring landslides through satellite imagery, thus contributing to more effective surveillance systems.

3 BACKGROUND

This section is divided into five parts, first, a brief introduction to the context of machine learning in Earth Observation (EO). Secondly, it delves into classical methods used for EO. Thirdly, it explores the advancements made in leveraging neural networks for EO tasks. Fourth, it details the methods and technologies employed in EO. Finally, it discusses the existing approaches and methodologies in the context of landslide risk management.

3.1 Introduction

Identifying a pattern or a feature on the surface is one of the primary objectives within the Earth Observation (EO) community. The emergence of cloud-based platforms has created new opportunities for addressing the issue through data-based algorithms that utilize the analysis-ready data (ARD) available on the Google Earth Engine (GEE) cloud computing platform[10].

Over the last two decades, a wide range of surface mapping methodologies have emerged within the EO domain. Broadly speaking, these methods can be classified into two main classes: knowledge-driven (Kd) and data-driven (Dd) approaches. Kd-based techniques emphasize logical underpinnings derived from physical principles, usually yielding more accurate results and favourable generalizations, albeit with considerable computational effort and identifiable scale limitations. In contrast, Dd strategies, employing large manually labelled datasets, have gained prominence in remote sensing due to the successful application of classical Machine Learning (ML) and Deep Learning (DL) techniques[13]. This proliferation of advanced surface mapping methodologies requires extensive prior knowledge and experience within the research community to effectively navigate and utilize these complex analytical tools.

3.2 Conventional machine learning techniques

As an analytical tool, one of the primary objectives and tasks of machine learning is to construct a model to represent complex, unknown, or incompletely understood relationships between data and target variables[14]. There are slight variations in types of machine learning algorithms; and these can be divided into two main categories according to their purpose: supervised learning algorithms and unsupervised learning algorithms. Supervised learning refers to the construction of a model to connect known inputs and outputs, but with unknown connections. Consequently, output values for new data can be predicted based on the relationships learned from previously processed training data [15]. Supervised learning can be divided into classification and regression problems. In classification problems, the predicted output is a label or semantic class. For example, to identify potential landslides, classification problems would label each pixel in an image as "landslide" or "non-landslide"[16]. Regression problems aim to predict a continuous variable. Some of the prevalent supervised learning algorithms include logistic regression (LR)[17], decision trees (DT)[17], support vector machines (SVM)[18], Naive Bayes (NB)[19], and artificial neural networks (ANN)[17]. In contrast to supervised learning, unsupervised learning algorithms seek patterns in data that hasn't been labeled, encompassing clustering techniques such as hierarchical clustering[20] and K-means[21].

3.3 Deep learning methods for remote sensing

As a subdiscipline of machine learning, deep learning is an extension of ANN. Deep learning uses multi-level deep neural networks to extract features from the input by progressively processing them. The scale and complexity of the networks are the main differences between deep learning and traditional ANN.

There are different methodologies for addressing the problem of landslide identification; for this investigation, we will focus on semantic segmentation, understood as the most efficient for pixel classification problems. The objective of segmentation is the same as that of traditional remote sensing image classification, which is usually achieved using traditional machine learning techniques, such as Random forests[22] and maximum likelihood classification [23].

To implement segmentation using neural networks, a convolutional neural network (CNN) architecture is used, which is trained on a previously labelled dataset. The network consists of multiple layers of neurons, where each layer processes the input information and produces an output that is used as input for the next layer. In this case, the network receives an image as input and produces an output containing a segmentation mask for each pixel in the image. For each pixel of the image, a label corresponding to a specific region is assigned according to the probability calculated by the neural network, which in this case is considered landslide and non-landslide.

There are different convolutional architectures for semantic segmentation such as U-net[24], Mask R-CNN[25], Feature Pyramid Network (FPN) [26], SegNet[27], or DeepLab[28]. In this investigation, we will primarily focus on U-net[24], which is one of the most widely used image segmentation algorithms in remote sensing tasks.

The U-net model, originally developed for biomedical image segmentation, features an architecture that comprises an encoding network (contraction path) followed by a decoding network (expansion path). The contraction path, which can be based on various convolutional neural network architectures such as VGG[29], ResNet[30], DenseNet[31], and Inception[31], captures the image information and progressively reduces spatial resolution. Each block within the contraction path consists of two convolution layers followed by a pooling layer, allowing the encoding of input images into multi-level feature representations and facilitating feature extraction while pooling layers serve to reduce the resolution.

Conversely, the decoder constitutes the second half of the architecture with the aim of semantically projecting the discriminatory features (lower resolution) learned by the encoder into pixel space (higher resolution) to achieve dense classification. The integral part of the U-net model is composed of upsampling and concatenation followed by regular convolution operations.

Finally, SegFormer[32] is a framework which unifies Transformer[33] with lightweight multilayer perception (MLP) decoders. This architecture uses multi-head attention to capture feature information from different areas of the image and combine it to generate an accurate segmentation output. Moreover, SegFormer uses a technique called "iterative positional encoding" to encode the position information of the image and enhance segmentation accuracy.

3.4 Earth Observation Techniques

Recent advances in Big Data processing and analysis have catalyzed significant shifts in both science and the broader society [34]. These advancements have spurred growth in many machine learning domains, notably in the socio-environmental applications of Earth Observation (EO). In this domain, Big Data has established itself as a central hub, facilitating the cloud-based analysis of voluminous EO datasets. This capability has expanded the possibilities for exploring terrestrial phenomena using EO techniques, offering a more efficient and detailed perspective for monitoring and evaluating both natural resources and human-generated environmental conditions.[9]

Although the use of EO techniques for the manipulation and analysis of satellite imagery is now several decades old, the democratization of computational power through cloud processes has enabled an increase in large-scale research and applications. Platforms such as Sentinel Hub, Earth on AWS, Microsoft Planetary, Open Data Cube (ODC), System for Earth observation data access, Processing and Analysis for Land monitoring (SEPAL) and mainly Google Earth Engine (GEE)[10] have contributed greatly to this advancement. Gomes (2020) [9] provides a more exhaustive comparison of these platforms. Within these, it is important to highlight GEE for its pioneering implementation of global access to a database of different satellite products and for providing the computational capability for real-time processing. In addition, GEE facilitates the implementation of APIs in languages accessible to a large part of the scientific community, such as Python, JavaScript and R [35], and its applications allow integration with artificial intelligence techniques[8].

This research utilizes the Google Earth Engine (GEE) architecture, as depicted in Fig 1. The diagram demonstrates GEE's layered structure, from user interfaces like the Earth Engine Code Editor and third-party apps at the top, to data storage components at the base. Central to its functionality are on-the-fly and batch computation systems, all seamlessly interconnected. This framework not only facilitates storage, preparation, and real-time analysis of spatial data but also taps into the robust computational power of Google servers for efficient processing.
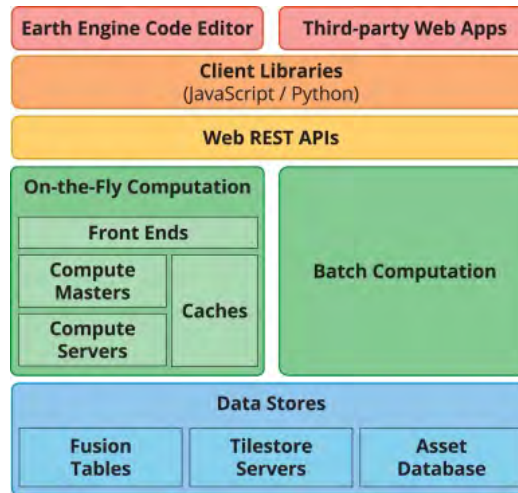
Figure 1: Google Earth Engine Architecture (Gorelick, 2017), EO tool used to access, analyze and publish open-access satellite products

3.5 Landslide prevention

A landslide, as defined by Highland and Bobrowsky[36], is a downward movement of rock, soil, or both, occurring along a rupture surface, a surface along which a displacement of soil material has occurred. Such landslides are not only common but also pose potentially catastrophic geological hazards[37]. To effectively mitigate their risks, disaster risk management literature emphasizes the importance of reliable monitoring, evaluation, and pinpointing vulnerable areas[38, 39]. Various studies investigate landslides triggered by earthquakes or heavy rainfall. For those induced by earthquakes, global databases[40, 41] capture both pre- and post-landslide events with fatal outcomes. More localized research primarily centers on susceptibility mapping, using methods ranging from deterministic and probabilistic models to heuristic and geostatistical techniques. This includes spatio-temporal analysis[42, 43] and assessments of anthropogenic impacts[44]. For rain-triggered landslides, NASA's repositories offer resources[45], utilizing the Global Precipitation Measurement (GPM) satellite constellation[46]. There are also studies on multitemporal analysis[47] and the specifics of rainfall-induced landslides[48].

Gomez (2023)[7] provides a comprehensive compilation of the four most significant global databases on landslides: The International Disaster Database (EM-DAT), the Disaster Inventory System (DesInventar), the Global Landslide Catalogue (GLC), and the Global Fatal Landslide Database (GFLD). These databases furnish critical landslide information for various regions worldwide. Using these sources, Gomez established the Unified Global Landslide Database (UGLD), encompassing 161 countries and documenting 37,946 landslides with 185,753 fatalities between 1903 and 2020. According to this compilation, regions with the highest frequency of such events are predominantly in Asian and Latin American countries. Notably, Peru is highlighted as the third country with the most landslides per 1000km2, only surpassed by El Salvador and Colombia. Moreover, Peru ranks among the countries with the most landslide-induced fatalities, as illustrated in figure 2.

As depicted in the figure, Peru is one of the countries most affected by landslides, yet there are relatively few studies focused on the identification and prevention of these phenomena in the country. This underscores the significance of the present research in contributing to landslide effect mitigation and serving as a valuable tool for decision-makers.

Furthermore, according to the statistical bulletin from the National Institute of Civil Defense of Peru (INDECI), between the years 2003 and 2021, considering phenomena such as avalanches, collapses, landslides, and flash floods (locally known as "huaycos"), there have been 8,347 recorded emergencies and 475 fatalities.[49] This highlights the critical need for focused research on the identification and prevention of landslides in this highly susceptible region.
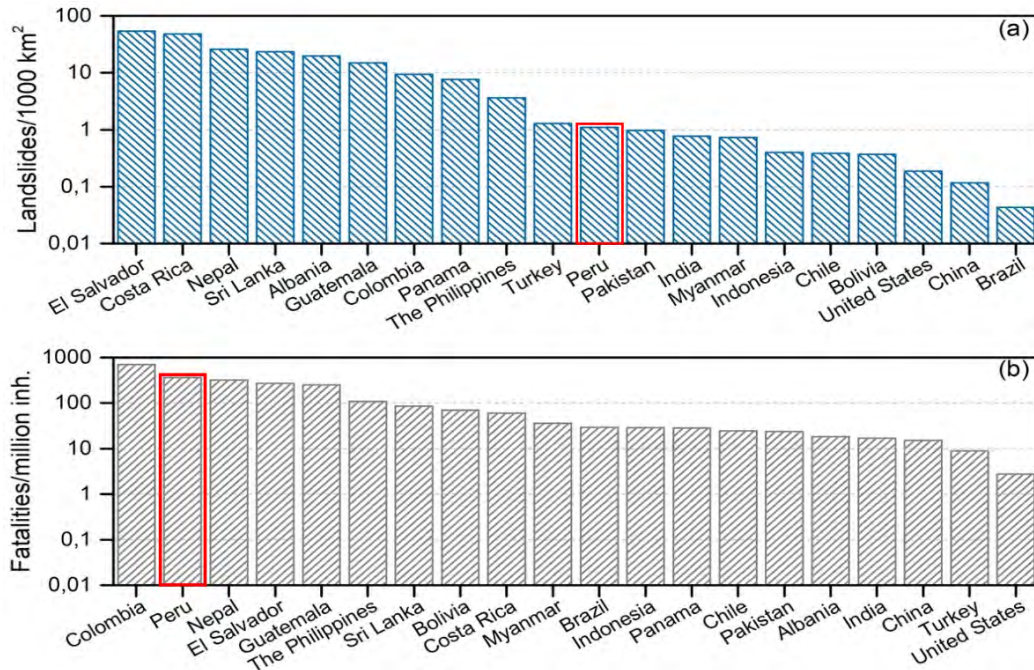
Figure 2: Distribution of the number of landslides and fatalities by country (Source: UGLD[7])

4 DATA AND RESOURCES

4.1 Landslide4Sense (L4S)

Landslide4Sense (L4S) [6] is a dataset consisting of non-georeferenced images of landslides from four distinct study areas: Iburi-Tobu in Japan, Karnataka in India, Bagmati in Nepal, and Taitung County in Taiwan. The dataset comprises 3,799 patches, each of which is a composite of 14 bands. These bands include all the spectral bands from Sentinel-2 [1] as well as elevation data from Alos Palsar satellite imagery [50]. Additionally, a slope band in degrees is calculated based on the elevation data.

Given that this dataset utilizes two different satellite products, it is necessary to harmonize the spatial resolutions of both products, thus requiring an adjustment to the pixel size. To this end, Table 1 shows the original spatial resolution of each band of the products, noting that there are bands of 10, 20, and 60 meters for Sentinel-2, and 12.5 meters for Alos Palsar. Furthermore, this dataset uses Sentinel-2 L1C product that considers the cirrus band (Band 10) and it does not appear in the L2A product, for the other hand, Alos Palsar is an image with an elevation band, and the slope band (degree) was calculated using the Horn algorithm[51] implemented in the Terra library[52] of the R programming language. The authors of this product perform a resize of all bands to a resolution of 10 meters per pixel to be subsequently used in pixel-wise segmentation tasks. These images have a size of 128x128 pixels and are properly labelled into two classes, landslide (1) and no landslide (0).

4.2 Own dataset (LS4-PE)

Drawing upon the L4S dataset, which is primarily situated within an Asian context, we expand this dataset to include a local context - the Peruvian forest. This results in a new dataset that complements the original one and maintains the same characteristics as those offered by L4S. It encapsulates the same 12 bands of the Sentinel-2 product and the elevation band from the Alos Palsar product.

In creating this new dataset, subsequently referred to as L4S-PE, we implement a pipeline based on CloudSen12[53]. This assists us in expanding a global dataset for segmentation with local characteristics. The process commences with identifying the location of the phenomena under evaluation (landslides) within a region of interest. This is followed by downloading the satellite images for these specific areas, applying semi-automatic labelling to these samples, and ultimately standardizing them for segmentation tasks.

---

[1]Band 8A was not considered due to its absence in the L4S dataset; furthermore, it does not provide additional relevant information for describing geomorphological changes.

6

Source	Band	Description	Original resolution (m)
Sentinel2	B01	Aerosol	60
Sentinel2	B02	Blue	10
Sentinel2	B03	Green	10
Sentinel2	B04	Red	10
Sentinel2	B05	VNIR 1	20
Sentinel2	B06	VNIR 2	20
Sentinel2	B07	VNIR 3	20
Sentinel2	B08	NIR	10
Sentinel2	B09	Vapor	60
Sentinel2	B10	Cirrus	60
Sentinel2	B11	SWIR 1	20
Sentinel2	B12	SWIR2	20
Alos Palsar	B13	Slope	12.5
Alos Palsar	B14	Elevation	12.5

Table 1: Set of bands used in the L4S and L4S-PE databases from Sentinel-2 and Alos Palsar satellite images. B08A band from S2 was not included

4.2.1 Study area

While the regions in Peru with the highest number of landslides are typically found between the coastal areas and the mountain range, according to the NASA Global Susceptibility Map[54] and the Peruvian Landslide Susceptibility Map (INGEMMET)[55], as well as the geographical characteristics similar to those of the regions in the L4S database, this study mainly focuses on the regions between the departments of Junin, Pasco, and Huánuco. These regions are located to the east of the Andes in Peru, at the confluence of the highlands and the jungle, transitioning from rugged to flat terrain. The study areas were defined based on their geomorphological characteristics, where the terrain is primarily covered by forests. This dense vegetation facilitates visual interpretation of optical remote sensing images, as shown in Figure 3i.

Although the regions in Peru that exhibit the highest incidence of landslides are predominantly situated between coastal and mountainous areas due to human impact, an analysis leveraging both NASA's Global Landslide Susceptibility Map (referenced as [54]) and the Peru Mass Movements Susceptibility Map by INGEMMET (cited as [55]) has been conducted. This research primarily focuses on the regions spanning the departments of Junin, Pasco, and Huánuco. These locales are strategically positioned to the east of the Andean range in Peru, marking the transitional zone between the highlands and the jungle, and shifting from rugged to flat topographies. The areas under study have been meticulously selected based on their distinct geomorphological characteristics, with a predominant coverage of forested terrain. This lush vegetative cloak significantly enhances the efficacy of visual interpretation applied to optical remote sensing imagery, as depicted in Figure 3i. This alignment with the geographical traits of the L4S database regions further solidifies the relevance and applicability of this study.

4.2.2 Image patches selection

We selected 1,000 regions of interest (ROIs) spread throughout the study area to obtain raw L4S-PE data. Each ROI measures 1500 x 1500 meters. In choosing each ROI, we ensured the presence of landslide events by visually inspecting the Google Earth platform, which provides very high-resolution images (sourced from CNES/Airbus, WorldView, Ikonos, etc.), and events that took place before 2018. For every ROI, we downloaded all Sentinel-2 L1C bands from the GEE datasets and the DEM band from Alos Palsar via the Alaska Satellite Facility (ASF) utilizing the RGEE package[35]. In the end, despite inherent resolution differences among bands, we harmonized all of them to a 10-meter spatial resolution, consistent with the approach adopted in the L4S product.

Then, the database was trimmed down to 838 patches after conducting a visual quality control check of the downloaded images. ROIs with cloud cover that hindered landslide identification were removed. Similarly, we excluded patches where sensors exhibited noise and areas where the identified feature was not a landslide but
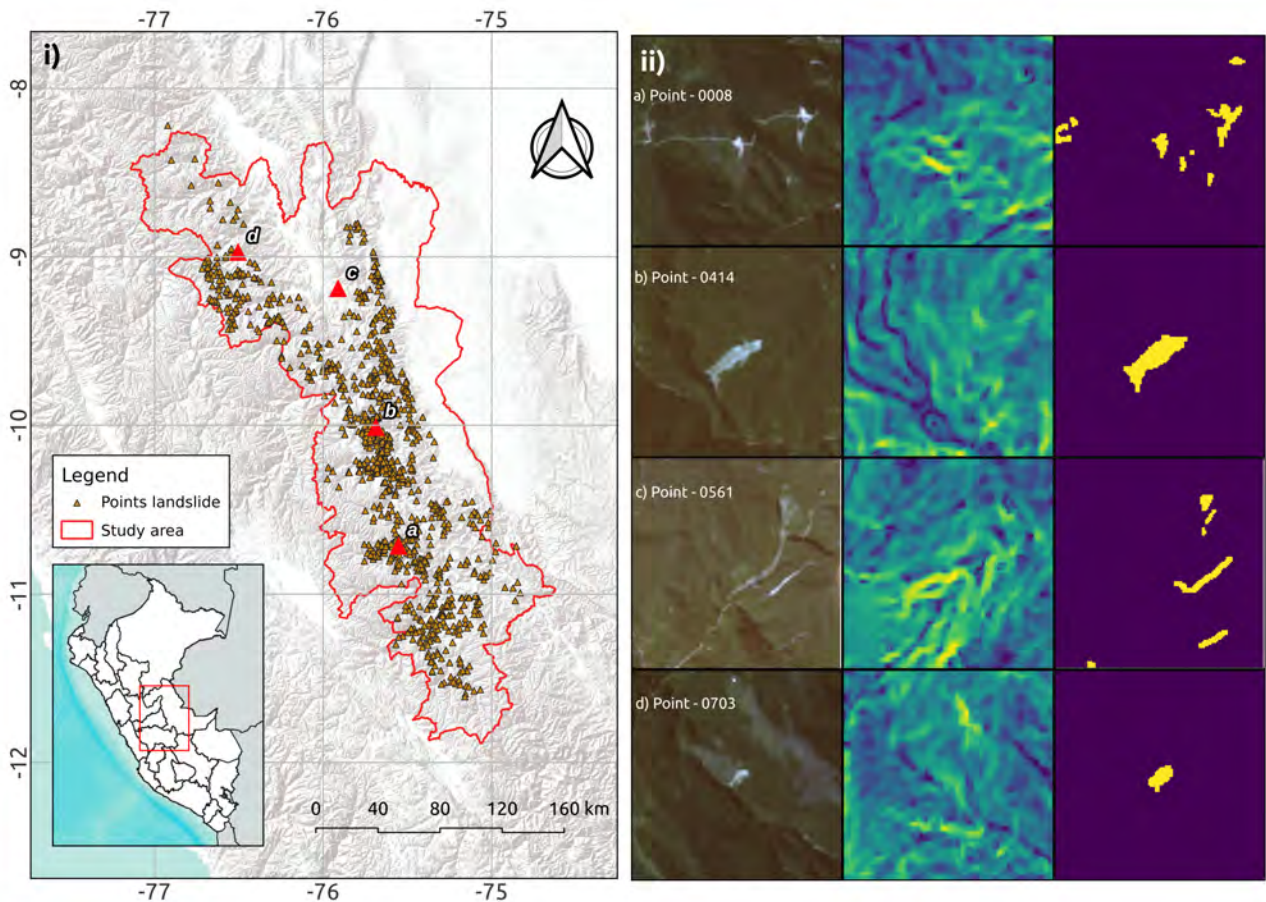
Figure 3: (i) shows the study area and ROI collection for L4S-PE. (ii) shows 4 examples of ROIs, RGB combination, slope band and labelled image from left to right

rather another type of land cover, such as agriculture. Figure 4 shows the download process and data source, as well as the number of patches assessed.

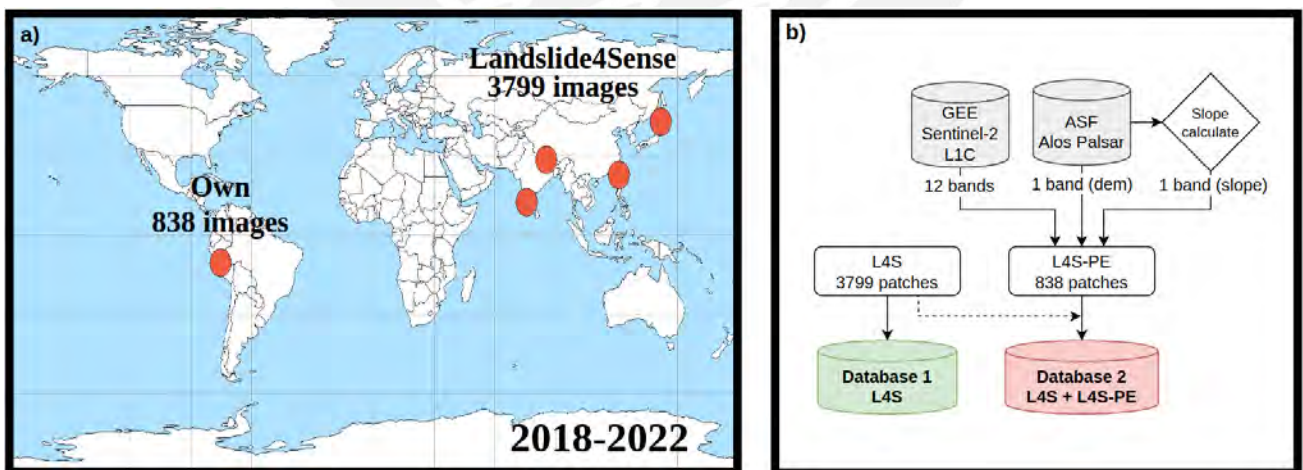Figure 4: (a) shows the data source, emphasizing the own database (L4S-PE). (b) shows the downloading process and integration to create the two databases evaluated

4.2.3 Labelling process

For the annotation of these images, we employed a modified version of the Intelligence foR Image Segmentation (Iris) active learning software[56], which facilitates manual labelling for each individual image while also enabling

the analysis of diverse band combinations (Figure 5). Each ROI possesses dimensions of 1500 x 1500 meters, thereby offering a spatial context for the target area measuring 1280 x 1280 meters. Upon completion of the annotation process, the labelled patch outcomes were exported in Hierarchical Data Format (HDF) to facilitate subsequent integration with machine learning techniques.

Figure 3ii shows three points from the database with their respective band combination, RGB (B4, B3, B2), slope (B13) and its mask manually created. Landslide in yellow.
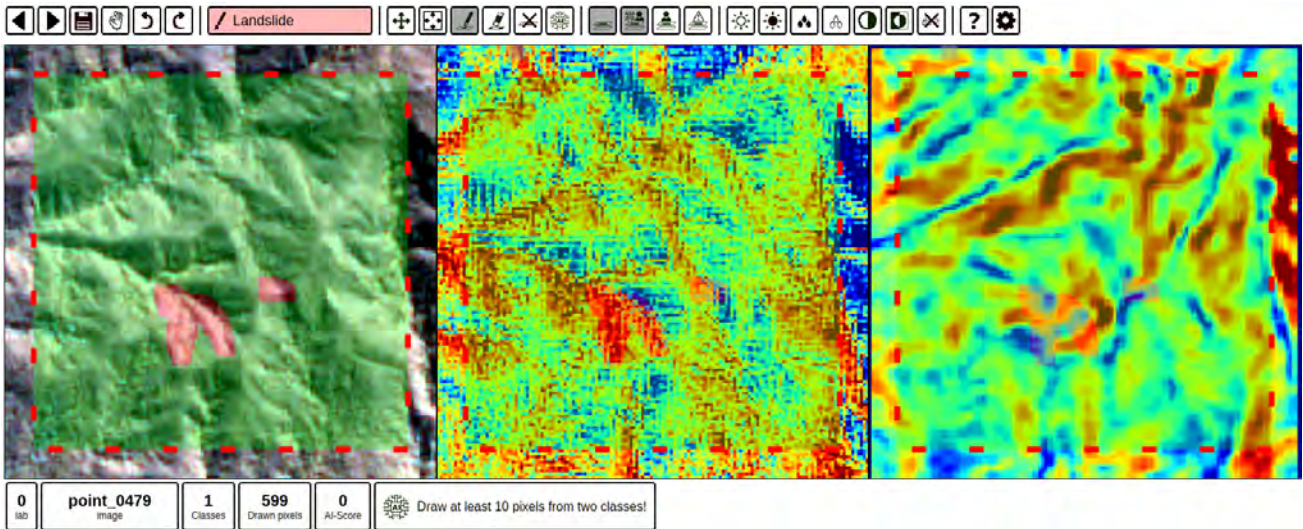

Figure 5: Iris interface for manual labelling, in this case, there are 3 images synchronized showing RGB, slope and elevation combination. The red dashed box shows the area of analysis of 128x128 pixels.

4.3 Database comparison

This study conducts experiments utilizing two sets of databases: the original Landslide4Sense (L4S) dataset and the uniquely compiled Landslide4Sense Peru (L4S-PE) dataset. Data normalization was carried out employing a min-max scaling, which transforms the behavior of the Sentinel-2 wavelength bands, as well as the elevation and slope values obtained from Alos Palsar, to a standardized range of 0 to 1.

A comparative analysis was conducted between the two data sets on a band-by-band basis, assessing the pixel distribution for each band. Figure 6 illustrates the data behavior after applying min-max normalization for each data source across all analyzed bands. Given that the wavelength ranges are consistent between the sets, one would expect similar data behaviors, except for local variations arising from topographical conditions or specific land cover types captured in the images. A notable difference is observed in Band 1 (the Aerosol Band), but this discrepancy does not preclude the possibility of using a unified database for modeling purposes.

Meanwhile, the table 2 shows the number of pixels identified as slips in both databases, evidencing how unbalanced it is.

Source	Landslide area (km^2)	% Landslide / No landslide	Number of patches
L4S	137.56	2.21 / 97.79%	3799
L4S-PE (own)	24.71	1.8 / 98.2%	838
L4S + L4S-PE	162.27	2.14 / 97.86%	4637

Table 2: Distribution of classes by database, area in square metres of landslides and number of patches for each database. Each patch has 128x128 pixels and 10 metres of spatial resolution.

5 METHODOLOGY

In this study, the methodology compared the efficiency of various deep learning models for semantic segmentation, primarily based on the Unet framework, with different architectures serving as encoders. Experiments were conducted employing two distinct databases: the original L4S and a combined set comprising L4S and
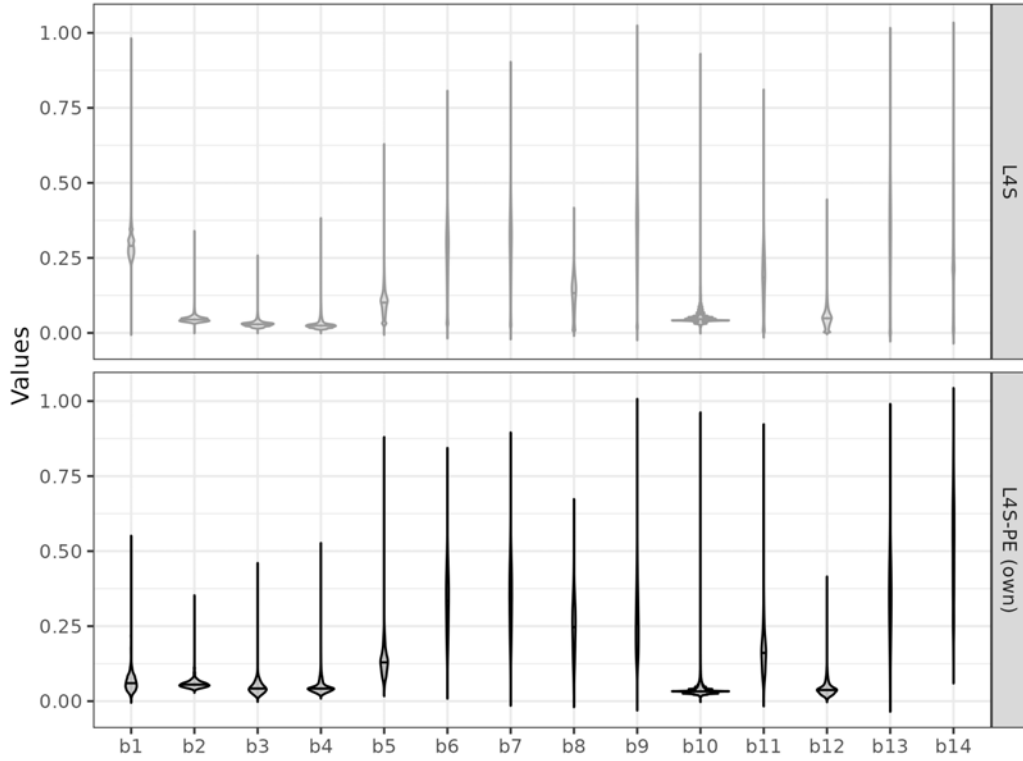
Figure 6: Comparative analysis evaluating distribution and variability for each spectral band values across the datasets L4S and L4S-PE (own)

L4S-PE. Two band combinations were evaluated: one incorporating all 14 bands present in the databases and another focusing on the six bands that originate with an initial resolution of 10 meters (R, G, B, NIR, Slope, and Elevation bands). These databases were partitioned into training and validation subsets. Furthermore, a variety of metrics were utilized to assess the performance efficacy of the machine-learning models under investigation.

5.1 Loss functions for segmentation

The segmentation algorithms tested in this research performed the same task of binary classification at the pixel level. This predicts the label of each pixel in 128x128 size images to classify whether it belongs to a landslide region or not.

5.1.1 Binary Cross-Entropy (BCE)

Binary Cross Entropy loss (BCE) per pixel serves as a fundamental measure frequently employed in image segmentation tasks within the machine learning and deep learning domains. For landslide extraction, there are only a landslide class and a non-landslide class. The BCE loss quantifies the divergence between the predicted and true probability distributions for each pixel, as defined in Equation 1:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^{n} (Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log(1 - \hat{Y}_i)) \tag{1}$$

In this equation, $Y_i \in 0, 1$ is a binary class label for pixels i with 1 being positive and 0 being negative. While $\hat{Y}_i \in [0, 1]$ refers to the predicted probability of the pixel being classified in the positive class, $(1 - \hat{p}_i \in [0, 1])$ is the probability of the pixel being classified as the negative class. The total number of pixels in the input image is indicated by the letter N.

5.1.2 Weighted Cross-Entropy (WCE)

Binary Cross-Entropy (BCE) loss function can be challenging to optimize due to the significant class imbalance between landslide and non-landslide pixels. A model trained with BCE could achieve high accuracy by pre-

dominantly classifying pixels as non-landslide, which would be misleading and ineffective for actual landslide extraction. To address this, the Weighted Cross-Entropy (WCE) loss is employed to account for the class imbalance and improve the model's predictive performance. Unlike the BCE loss, WCE adds weight to the positive examples, i.e. by assigning smaller weights to the dominant background weights and larger weights to the foreground weights. The equation is as follows:

$$\mathcal{L}_{WCE} = -\frac{1}{N} \sum_{i=1}^{n} (\beta Y_i \log \hat{Y}_i + (1 - Y_i) \log(1 - \hat{Y}_i)) \tag{2}$$

Where β is the coefficient that weights the loss of positive examples. Whereas setting $\beta > 1$ reduces false negatives and consequently increases recall, $\beta < 1$ cuts false positives, thus increasing precision.

5.1.3 Dice Loss

The Dice index, commonly used in segmentation tasks[57], serves as a measure of the overlap between two sets. It measures the overlap between the predicted and target segmentation masks and provides a differentiable and smooth measure of segmentation accuracy. It is particularly effective when dealing with imbalanced datasets and when the focus is on capturing fine details in the segmentation masks. The equation is defined as follows:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{3}$$

In this equation, X and Y represent the predicted and ground-truth sets of pixels, respectively. $|X \cap Y|$ is the size of the intersection of the two sets, and $|X|$ and $|Y|$ are the sizes of the predicted and ground-truth sets, respectively. The variable N, which denotes the total number of pixels in the input image, usually has a dimension of 128 x 128 in our study.

5.1.4 WCE-Dice loss

Combining multiple loss functions is a viable approach to optimize the model's performance. In this case, we used the WCE-Dice loss, which is a weighted sum of both Weighted Cross-Entropy (WCE) and Dice losses.

$$\mathcal{L}_{WCE-Dice} = (1 - \alpha)\mathcal{L}_{WCE} + \alpha\mathcal{L}_{Dice} \tag{4}$$

Where α is the coefficient that weights the Dice loss $\mathcal{L}_{Dice}$, against the WCE loss, $\mathcal{L}_{WCE}$.

5.2 Evaluation metrics

For the quantitative evaluation of landslide detection and segmentation performance, the following metrics were used: precision, recall, and F1-score. Precision allows identifying how many landslide areas were correctly detected; recall is used to determine how many of the landslides in the image were detected; and the F1-score provides approximately the mean of the two values when they are close and is generally higher than the harmonic mean. These metrics were calculated based on true positives (TP), false positives (FP), and false negatives (FN).

It is worth mentioning that, when dealing with a naturally imbalanced dataset, the accuracy metric is not well-suited to this type of problem mainly because it does not consider the imbalance in class distribution.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = 2\text{x} \frac{Precision \cdot Recall}{Precision + Recall} \tag{7}$$

$$IoU(JaccardIndex) = \frac{TP}{TP + FP + FN} \tag{8}$$

5.3 Experimentation and results

In this section, we present a series of experiments conducted on the landslide datasets described in the preceding section. These experiments serve to evaluate various machine learning models and their performance using the metrics previously outlined.

5.3.1 Computational environment

The experiments were implemented in the Python programming language using the following characteristics:

Configuration	Content
Operative System	Ubuntu 22.04LTS
GPU	NVIDIA RTX 3070 (16 GB RAM)
Deep-Learning Framework	Pytorch and Pytorch Ligthning
Parallel Computing Platform	CUDA 12.1
Programming Language	Python 3.8.10

Table 3: Computational environment used for experiments

5.3.2 Implementation details

The optical remote sensing images utilized in this study were sourced from Google Earth Engine (GEE). From GEE's data mosaics, we acquired images from Sentinel-2 and Alos Palsar satellites. The spectral characteristics, image resolutions, and temporal range of the selected imagery were carefully assessed to ensure consistency. Each patch in the dataset measures 128x128 pixels, encompasses 14 spectral bands, and features a 10m x 10m spatial resolution, sourced from images spanning from 2018 to 2022. In total, the study comprises a labeled image collection of 4637 patches, of which 3799 belong to the original L4S dataset and 838 to the L4S-PE dataset.

In this research, landslide identification was formulated as a semantic segmentation task. The labeled images were converted to a format compatible with the segmentation algorithms employed in the study. We conducted multiple experiments, varying both the data source (either L4S or L4S combined with L4S-PE) and the set of spectral bands utilized. Specifically, we considered two sets of bands: the original 14-band set and a subset featuring bands with an original resolution of 10 meters (i.e., B2, B3, B4, B8 from Sentinel-2, and slope and elevation from Alos Palsar). Given these conditions, the dataset was divided into four experimental groups, as detailed in Table 4. In all cases, the datasets were partitioned into training and validation subsets, following an 80%-20% split. To ensure reproducibility, a fixed random seed was used for the dataset splitting and batch generation processes.

Source	Dataset	n patches	Bands Exp 1	Bands Exp 2
L4S	Training	2661	128 x 128 x 14	128 x 128 x 6
L4S + L4S-PE	Training	584	128 x 128 x 14	128 x 128 x 6
L4S	Validation	1138	128 x 128 x 14	128 x 128 x 6
L4S + L4S-PE	Validation	254	128 x 128 x 14	128 x 128 x 6

Table 4: Database specifications for databases used in the experiments

5.3.3 Experiments

The L4S and L4S-PE databases were employed to evaluate the models for landslide segmentation. Figure 7 displays sample images from these databases along with their corresponding labeled images to illustrate the relative proportions of landslide(1) and non-landslide(0) areas.

Experimentation was conducted using the PyTorch Lightning framework, starting with the configuration of a dataloader for reading and normalizing the datasets. A data loader was configured for reading local h5-formatted databases sourced from Google Earth Engine (GEE). The experiments were categorized based on the spectral bands used: one group utilized all 14 bands, while the other focused on bands with an original 10-meter resolution as per Table 1. Min-max normalization in the range of 0 to 1 was applied to each image.
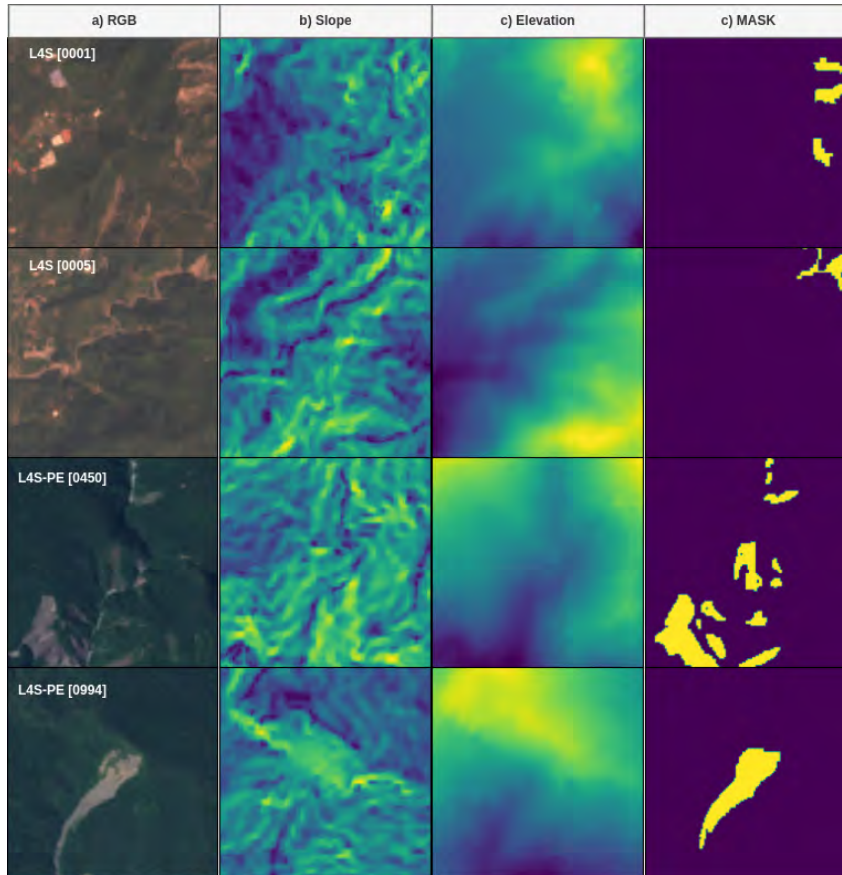
Figure 7: Images from two datasets (L4S and L4S-PE) where a) RGB corresponds to bands B4, B3, B2 b) Slope corresponds to B12, c) Elevation corresponds to B13 and d) Mask is the image labeled in landslide and no-landslide

The datasets were divided into training and validation sets at an 80/10% ratio. All experiments utilized the U-net architecture, taking into account memory requirements for both encoder and decoder components, and included large tensor skip connections. The Residual Network (Resnet34) and MobilenetV2 served as backbones for these experiments. Memory requirements were calculated considering batches of 16 images with dimensions $16x[14,6]x128x128$, using the Adam optimizer with a learning rate of 10^{-3} and the WCE-Dice loss function. During training, the learning rate was reduced by a factor of 0.1 every 30 epochs. Given the data imbalance, a 5-to-1 weight was assigned in favor of the landslide class over the background. If no improvement in the loss function was observed after 10 epochs, the model with the lowest validation loss was selected. Details of these configurations are provided in Table 5.

Backbone	Skip connection	Pooling	Codif. levels	N° Parameters
Unet (Vanilla)	Direct connection	Max-pooling	5	$\sim 23{,}4M$
Resnet34	Residual block	Global Average Pooling	5	$\sim 24{,}4M$
MobilenetV2	Residual block	Global Average Pooling	6	$\sim 6{,}6M$
Segformer	Multi-head Attention	Self-Attention	5	$\sim 16{,}4M$

Table 5: Models used for Unet encoders

5.3.4 Results

In this research, various Deep Learning models based on Unet and Transformers (Table 5) were evaluated for the task of semantic segmentation in satellite imagery, specifically from Sentinel2 and Alos Palsar sources. The models were trained to classify each pixel as either being part of a landslide or not. Furthermore, the impact of different normalization techniques and the number of channels utilized were also assessed. Detailed metrics for these experiments can be found in Table 6.

In the study focused on satellite image segmentation for landslide detection, the Unet (Vanilla) architecture

		Metrics				Loss function		
		F1-score	**Precision**	**Recall**	**IoU**	**Dice-WCE**	**WCE**	**Dice**
6 bands	Unet (Vanilla)	**0.747**	0.761	0.739	0.600	**0.223**	0.159	0.287
	Resnet34	0.731	0.784	0.691	0.581	0.238	0.177	0.298
	MobilenetV2	0.718	0.687	0.758	0.565	0.256	0.202	0.309
	Segformer	0.702	0.750	0.672	0.546	0.261	0.196	0.326
14 bands	Unet (Vanilla)	**0.755**	0.781	0.735	0.611	**0.215**	0.157	0.273
	Resnet34	0.734	0.785	0.695	0.585	0.233	0.174	0.292
	MobilenetV2	0.723	0.735	0.716	0.571	0.246	0.184	0.308
	Segformer	0.714	0.743	0.694	0.559	0.259	0.208	0.310

Table 6: Performance metrics of evaluated models on the validation set of the L4S database

		Metrics				Loss function		
		F1-score	**Precision**	**Recall**	**IoU**	**Dice-WCE**	**WCE**	**Dice**
6 bands	Unet (Vanilla)	**0.714**	0.749	0.687	0.559	**0.256**	0.191	0.321
	Resnet34	0.695	0.742	0.659	0.537	0.278	0.223	0.333
	MobilenetV2	0.688	0.751	0.640	0.528	0.286	0.237	0.335
	Segformer	0.669	0.712	0.639	0.507	0.299	0.245	0.353
14 bands	Unet (Vanilla)	**0.719**	0.744	0.701	0.566	**0.251**	0.186	0.317
	Resnet34	0.705	0.737	0.681	0.548	0.265	0.202	0.327
	MobilenetV2	0.693	0.747	0.654	0.535	0.277	0.216	0.339
	Segformer	0.681	0.727	0.648	0.521	0.288	0.230	0.347

Table 7: Performance metrics of evaluated models on the combined validation set of the L4S and L4S-PE databases

consistently outperformed more complex models like Resnet34, MobilenetV2, and Segformer. This superior performance was observed across two different datasets—L4S and a combined set of L4S and L4S-PE—as well as across configurations with 6 and 14 bands. The addition of more bands did lead to marginal improvements in F1-scores, most noticeably for the Unet model, suggesting that increased spectral information might be beneficial. However, this improvement was not universal across all architectures. Despite the relatively high F1-scores and Precision metrics, all models exhibited substantially lower IoU scores, pointing to a room for improvement in capturing the class overlap effectively. The trend of lower IoU was consistent when the data set was switched from L4S to the more complex combined dataset, suggesting challenges related to additional complexity or variance in the latter. Furthermore, the Dice-WCE loss function tended to yield lower values compared to using Dice or WCE alone, indicating its potential effectiveness in balancing the competing objectives of this specific segmentation task.

In Figure 8, we graphically present the segmentation outcomes across various models. The results indicate that the segmentation performance is relatively stable, irrespective of the changes in the dataset, employed model, and the number of spectral bands used. This stability is advantageous when the task involves binary classification and does not necessitate a complex model architecture for feature extraction from the different band sets. Moreover, landslide identification might become more challenging when shifting to a different geographic region, as the employed datasets are tailored to specific geographic features.

Landslide4Sense (L4S) itself highlights the efficacy of 12 different models, with SQNet and ResUnet emerging as the most effective, registering F1-scores of 70.24% and 71.65%, respectively. These findings affirm that these architectures excel in landslide detection, a success attributed to their balanced information flow between shallow and deep network layers. In this research, we achieved an F1-score of 0.755 using a Vanilla Unet model combined with 14 spectral bands and the L4S database. The noteworthy performance is mainly ascribed to the adoption of a combined Weighted Cross-Entropy (WCE) and Dice Loss function, which is particularly suitable for handling unbalanced data.

While the employed datasets in this study are tailored to specific geographic regions, providing robust models for those areas, the external validity of these models to other geographic contexts remains an open question.
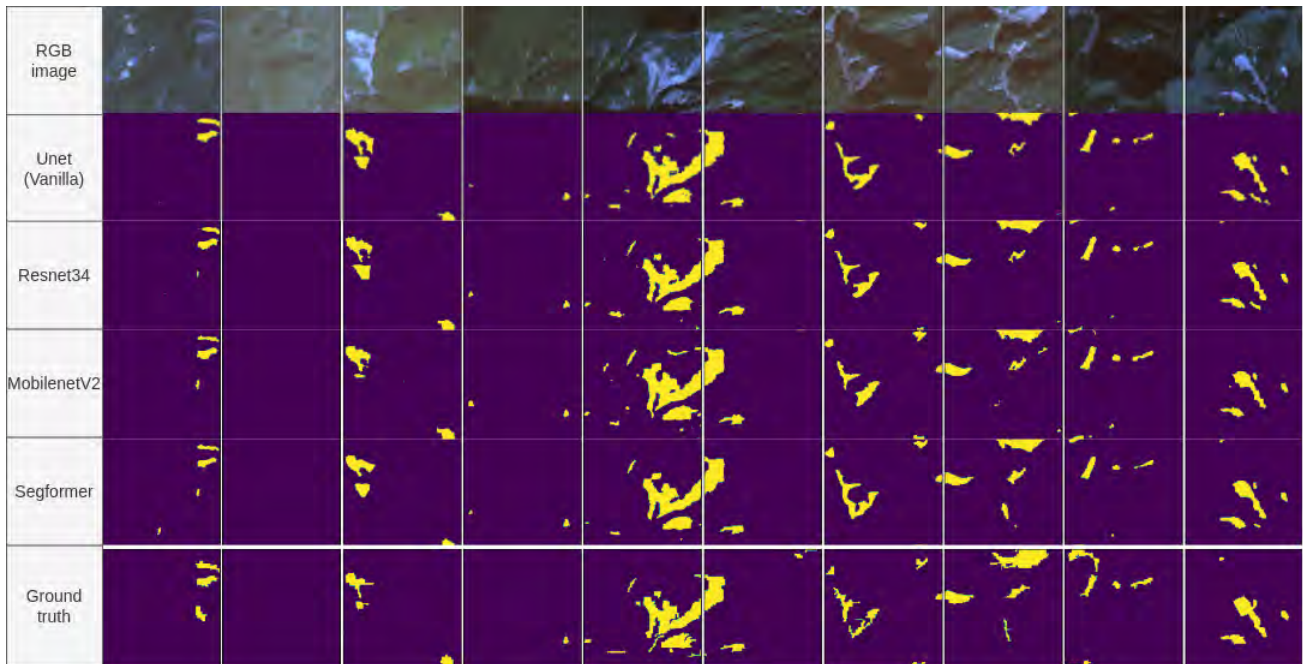
Figure 8: Comparative performance of top-performing models in identifying Landslide and Non-Landslide areas, featuring examples of RGB patches and corresponding ground truth

Adaptability to varying geographic features—such as soil type, vegetation cover, and topography—could mean additional layers of complexity in the model or fine-tuning the existing architectures with locally-sourced, high-quality data. Moreover, temporal variations introduce another layer of complexity. Factors such as seasonal changes, which could influence soil moisture and therefore landslide propensity, are not accounted for in static datasets. Future research could explore integrating time-series satellite data to capture these temporal dynamics, potentially employing recurrent neural network (RNN) layers or Long Short-Term Memory (LSTM) units in the architecture to model these temporal dependencies. This temporal contextualization could enrich the model's predictive power and contribute to a more dynamic and adaptable landslide detection system.

6   CLOUD-BASED MONITORING SYSTEM FRAMEWORK

Remote sensing techniques have proven to be instrumental in monitoring various geographic and environmental phenomena, with applications spanning diverse regions globally [58, 59, 60, 61]. One of the most commonly used platforms for monitoring land cover changes is GEE, which provides a comprehensive environment that simplifies preprocessing, streamlines satellite image acquisition, enhances monitoring capabilities through AI techniques like classification and segmentation and offers cloud integration facilities.

This section is divided into two parts: the first introduces a temporal dimension to landslide assessment, while the second outlines the data flow process for integration into a monitoring system.

6.1   Temporal analysis

The focus of the current study is the resolution of a task within the domain of computer vision. Nonetheless, the efficacy of a monitoring tool is contingent upon temporal variables. For the landslide detection task of this research, we use Sentinel-2 imagery providing new observations every 5 days for a targeted area and limited with the cloud cover which can impede the temporal resolution of image acquisition [62].

On the other hand, Figure 9 elucidates the temporal analysis of imagery from different dates over an area of interest, revealing the alterations in surface coverage in areas that have experienced landslides. For each event, pre- and post-event imagery are displayed, thereby integrating a temporal factor into the satellite-based assessment and detection of landslide occurrences. The model used for these predictions were U-net Vanilla with 6 bands and using the full dataset (L4S + L4S-PE) trained previously (Table 7).

15

Figure 9: Comparative temporal analysis in a Region of Interest (ROI) using Sentinel-2 images for a landslide occurred in Huanuco, 2020[63]

6.2 Monitoring system

The adoption of remote sensing techniques for the monitoring of geographic and environmental phenomena has witnessed an exponential increase internationally [58, 59, 60, 61]. Geographical Information Systems (GIS) have progressively evolved to include cloud-based analyses of satellite imagery. The current investigation leverages the capabilities of Google Earth Engine (GEE), a pioneering cloud-based platform that streamlines cloud service integration, simplifies the download and preprocessing of imagery, and underpins AI-driven segmentation techniques for more sophisticated monitoring endeavors.

6.2.1 Download and imagery processing

Satellite imagery provides a unique view of the terrain and allows early detection of potential landslides. This study works with GEE, which allows filtering and selection tools based on parameters like dates, cloud cover and localization to download Sentinel-2 (L1C) images with a spatial resolution of 10 metres which are available every 5 days for the same location, limited only to the cloud cover of the scene at the time they were captured. On the other hand, Alos Palsar images are not directly in GEE but can be stored in a cloud bucked like Google Cloud Storage (GCS) in Geotiff format.

Working with the Pytorch Lightning framework, it facilitates the dataloader and splitting for an area of interest (aoi) to monitor and links to the pre-trained model to create a map of landslide predictions.

In cases when the area is bigger than the original dimensions, we implemented a stitching process that resembles the prediction in the same size of the aoi evaluated. This process exports the results in a binary raster with Cloud Optimized Geotiff (COG) format that facilitates the integration with viewers.

Figure 10 shows the workflow used in this research considering the main parts, data processing, model training and, analysis and application.

6.2.2 Cloud-Based integration

To establish a dynamic and scalable landslide monitoring system, the trained semantic segmentation models—designed to differentiate between landslide and non-landslide classes—are securely housed within Google Cloud Storage. This allows for robust and accessible data management. Similarly, other studies have utilized cloud-based architecture and on-the-fly processes with Google services, demonstrating the effectiveness of these approaches in environmental analysis and near real-time monitoring [64]. The backend infrastructure, potentially developed using Google Cloud Functions or analogous cloud services, acts as an intermediary between user interactions or system-driven triggers and the execution of the model's processes. This serverless architecture affords an economical scalability model, engaging resources solely during model inquiries and thus adeptly managing fluctuating workloads without persistent server upkeep.

The automation of the landslide detection workflow leverages satellite data acquisition and processing, utilizing scheduling services to periodically invoke GEE for the latest Sentinel-2 and Alos Palsar imagery about predefined areas of interest.

Upon the detection of new imagery, the pipeline activates a sequence of prearranged cloud functions to preprocess the data, ensuring compatibility with our deep learning models. The images, once processed, are directed through the semantic segmentation models to ascertain potential landslide areas. This system is not only streamlined for efficiency but also strategically reduces the necessity for manual oversight, thereby expediting the detection process.
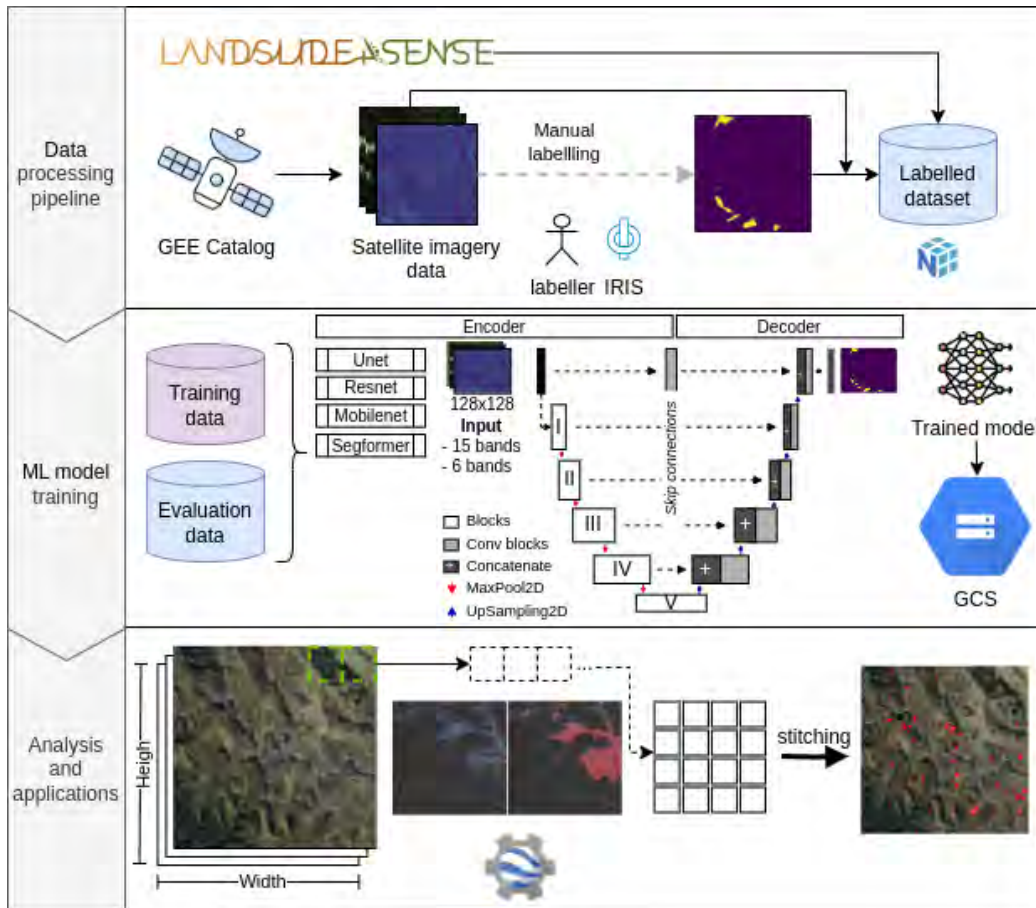
Figure 10: Workflow used of the Landslide Detection System using Sentinel-2 and Alos Palsar Imagery

It is envisaged that the landslide predictions formulated by the models will be systematically archived in Google Cloud Storage. In doing so, we guarantee that the predictive data remains promptly accessible for in-depth analysis and is readily amenable to integration with web-based visualization interfaces utilizing the Cloud Optimized Geotiff (COG) standard. Consequently, this would furnish end users, such as disaster response teams, with the capabilities to swiftly visualize and react to landslide incidents.

7 CONCLUSIONS

In the field of remote sensing, databases such as Landslide4Sense, which contain pixel-accurately labeled satellite imagery sourced from Sentinel-2 and Alos Palsar, exemplify the power of data-driven solutions. This research highlights that the fusion of additional data can amplify the scalability and replicability of machine learning models, making them applicable across varied geographical terrains.

As part of the research, we created a local satellite imagery dataset targeting landslides in the central regions of Peru with 838 images, referencing the foundational Landslide4Sense (L4S) dataset from Asian territories, comprising 3799 meticulously annotated images. These additions are congruent with the original dataset in terms of attributes and configuration.

Regarding the evaluation of models, we trained various models based on U-net and transformer architectures and proposed a synergistic combination of Dice Loss and Weighted Cross-Entropy (WCE) loss functions. This approach achieved an F1-score of 75.5% using the L4S database and 71.9% with the combined L4S and L4S-PE datasets, using the U-net (vanilla) architecture, both surpassing the benchmark dataset's score of 71.6%. It was found that the effectiveness of the fused loss function is particularly potent for imbalanced data. However, adaptations to geographic conditions and temporal changes, such as seasonal transitions, influence the accuracy of the models. This underscores the need for continuous expansion and refinement of datasets to boost the precision and performance of these models.

Finally, a comprehensive process has been developed that encompasses data acquisition, processing, model training/validation, and the evaluation of areas over different time ranges. Evidence of a change in the landslide surface itself was observed, with the potential to drive the general applicability of segmentation systems in landslide monitoring. This approach could broadly benefit the academic community and government stakeholders in Latin America and worldwide to enhance landslide prediction capabilities, utilizing advanced remote sensing and machine learning techniques to develop and implement effective preventive measures.

8 CODE AVAILABILITY

The code required for database generation, associated resource downloads for each ROI, annotation creation, deep learning model training, application construction, and publishing is open-source and available on Github at https://github.com/ryali93/lanslide4sat_pl. Additionally, the evaluated metrics from all conducted experiments are stored and can be accessed via the Weight & Biases tool [65], hosted on the following web repository: https://wandb.ai/ryali/lanslide4sat_pl.

9 ACKNOWLEDGMENTS

References

[1] Xuanmei Fan, Gianvito Scaringi, Oliver Korup, A. Joshua West, Cees J. van Westen, Hakan Tanyas, Niels Hovius, Tristram C. Hales, Randall W. Jibson, Kate E. Allstadt, Limin Zhang, Stephen G. Evans, Chong Xu, Gen Li, Xiangjun Pei, Qiang Xu, and Runqiu Huang. Earthquake-induced chains of geologic hazards: Patterns, mechanisms, and impacts. *Reviews of Geophysics*, 57(2):421–503, 2019.

[2] Tazio Strozzi, Jan Klimeš, Holger Frey, Rafael Caduff, Christian Huggel, Urs Wegmüller, and Alejo Cochachin Rapre. Satellite sar interferometry for the improved assessment of the state of activity of landslides: A case study from the cordilleras of peru. *Remote Sensing of Environment*, 217:111–125, 2018.

[3] ERIK Eberhardt, JJ Clague, and D Stead. Landslide monitoring: The role of investigative monitoring to improve understanding and early warning of failure. *Landslides: types, mechanisms and modeling*, pages 222–234, 2012.

[4] Renata Pacheco Quevedo, Andrés Velastegui-Montoya, Néstor Montalván-Burbano, Fernando Morante-Carballo, Oliver Korup, and Camilo Daleles Rennó. Land use and land cover as a conditioning factor in landslide susceptibility: a literature review. *Landslides*, 20(5):967–982, May 2023.

[5] Qiangqiang Yuan, Huanfeng Shen, Tongwen Li, Zhiwei Li, Shuwen Li, Yun Jiang, Hongzhang Xu, Weiwei Tan, Qianqian Yang, Jiwen Wang, Jianhao Gao, and Liangpei Zhang. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241:111716, 2020.

[6] Omid Ghorbanzadeh, Yonghao Xu, Pedram Ghamisi, Michael Kopp, and David Kreil. Landslide4Sense: Reference Benchmark Data and Deep Learning Models for Landslide Detection, June 2022. arXiv:2206.00515 [cs, eess].

[7] Derly Gómez, Edwin F. García, and Edier Aristizábal. Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, Mar 2023.

[8] Liping Yang, Joshua Driscol, Sarigai Sarigai, Qiusheng Wu, Haifei Chen, and Christopher D. Lippitt. Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review. *Remote Sensing*, 14(14):3253, July 2022.

[9] Vitor C. F. Gomes, Gilberto R. Queiroz, and Karine R. Ferreira. An overview of platforms for big earth observation data management and analysis. *Remote Sensing*, 12(8), 2020.

[10] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017. Big Remotely Sensed Data: tools, applications and experiences.

[11] Machine learning in earth engine, 2023. Última actualización: 21 de agosto de 2023.

[12] Lillianne Thomas and Ryan Avery. Deep learning with tensorflow: Tutorials for earth science applications, 2023. Última actualización: 2023.

[13] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, December 2017.

[14] Anuj Karpatne and Stefan Liess. A guide to earth science data: Summary and research challenges. *Computing in Science Engineering*, 17(6):14–18, 2015.

[15] Danilo Bzdok, Martin Krzywinski, and Naomi Altman. Machine learning: supervised methods. *Nature Methods*, 15(1):5–6, Jan 2018.

[16] Zhengjing Ma, Gang Mei, and Francesco Piccialli. Machine learning for landslides prevention: a survey. *Neural Computing and Applications*, 33(17):10881–10907, September 2021.

[17] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007.

[18] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 144–152, New York, NY, USA, 1992. Association for Computing Machinery.

[19] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, USA, 2008.

[20] Gabor J. Szekely and Maria L. Rizzo. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. *Journal of Classification*, 22(2):151–183, Sep 2005.

[21] Zejin Ding, Jian Yu, and Yanqing Zhang. A new improved k-means algorithm with penalized term. In *2007 IEEE International Conference on Granular Computing (GRC 2007)*, pages 313–313, 2007.

[22] Mahesh Pal. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222, 2005.

[23] Pushpendra Singh Sisodia, Vivekanand Tiwari, and Anil Kumar. Analysis of supervised maximum likelihood classification for remote sensing image. In *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, pages 1–4, 2014.

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[25] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. pages 2980–2988, 2017.

[26] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. 2016.

[27] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2016.

[28] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis amp; Machine Intelligence*, 40(04):834–848, apr 2018.

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[31] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[32] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers, 2021.

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[34] Viktor Sebestyén, Tímea Czvetkó, and János Abonyi. The applicability of big data in climate change research: The importance of system of systems thinking. *Frontiers in Environmental Science*, 9, 2021.

[35] Cesar Aybar, Quisheng Wu, Lesly Bautista, Roy Yali, and Antony Barja. rgee: An r package for interacting with google earth engine. *Journal of Open Source Software*, 2020.

[36] L.M. Highland and Peter Bobrowsky. *The landslide handbook—A guide to understanding landslides*. 2008.

[37] Huiming Tang, Janusz Wasowski, and C. Hsein Juang. Geohazards in the three gorges reservoir area, china – lessons learned from decades of research. *Engineering Geology*, 261:105267, 2019.

[38] Mohammad Azarafza, Mehdi Azarafza, Haluk Akgün, Peter M. Atkinson, and Reza Derakhshani. Deep learning-based landslide susceptibility mapping. *Scientific Reports*, 11(1):24112, December 2021.

[39] Enrique A. Castellanos Abella and Cees J. Van Westen. Qualitative landslide susceptibility assessment by multicriteria analysis: A case study from san antonio del sur, guantánamo, cuba. *Geomorphology*, 94(3):453–466, 2008. GIS technology and models for assessing landslide hazard and risk.

[40] Melanie J. Froude and David N. Petley. Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18(8):2161–2181, August 2018.

[41] Robert G. Schmitt, Hakan Tanyas, Margaret A. Nowicki Jessee, Jing Zhu, Kimberley M. Biegel, Kate E. Allstadt, Randall W. Jibson, Eric M. Thompson, Cees J. van Westen, Hiroshi P. Sato, David J. Wald, Jonathan W. Godt, Tanya Gorum, Chen Xu, Ellen M. Rathje, and Keith L. Knudsen. An open repository of earthquake-triggered ground-failure inventories (ver 4.0, october 2022). U.S. Geological Survey data release collection, 2017. Version: 4.0.

[42] Tolga Görüm and Seçkin Fidan. Spatiotemporal variations of fatal landslides in turkey. *Landslides*, 18(5):1691–1705, May 2021.

[43] M. A. Nowicki Jessee, M. W. Hamburger, K. Allstadt, D. J. Wald, S. M. Robeson, H. Tanyas, M. Hearne, and E. M. Thompson. A global empirical model for near-real-time assessment of seismically induced landslides. *Journal of Geophysical Research: Earth Surface*, 123(8):1835–1859, 2018.

[44] Hakan Tanyaş, Kevin Hill, Luke Mahoney, Islam Fadel, and Luigi Lombardo. The world's second-largest, recorded landslide event: Lessons learnt from the landslides triggered during and after the 2018 mw 7.5 papua new guinea earthquake. *Engineering Geology*,

297:106504, 2022.

[45] Dalia Kirschbaum, Thomas Stanley, and Yaping Zhou. Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, 249:4–15, 2015. Geohazard Databases: Concepts, Development, Applications.

[46] Gail Skofronick-Jackson, Walter A. Petersen, Wesley Berg, Chris Kidd, Erich F. Stocker, Dalia B. Kirschbaum, Ramesh Kakar, Scott A. Braun, George J. Huffman, Toshio Iguchi, Pierre E. Kirstetter, Christian Kummerow, Robert Meneghini, Riko Oki, William S. Olson, Yukari N. Takayabu, Kinji Furukawa, and Thomas Wilheit. The global precipitation measurement (gpm) mission for science and society. *Bulletin of the American Meteorological Society*, 98(8):1679 – 1695, 2017.

[47] Kushanav Bhuyan, Hakan Tanyas, Lorenzo Nava, Silvia Puliero, Sansar Raj Meena, Mario Floris, Filippo Catani, Cees Van Westen, and Tanuj Pareek. Multi-spatiotemporal landslide mapping for landslide evolutionary investigation. pages EGU22–1356, May 2022.

[48] Pukar Amatya, Dalia Kirschbaum, and Thomas Stanley. Rainfall-induced landslide inventories for lower mekong based on planet imagery and a semi-automatic mapping method. *Geoscience Data Journal*, 9(2):315–327, 2022.

[49] Boletín estadístico virtual de la gestión reactiva del indeci - 2023. Virtual, 2 2023.

[50] JAXA/METI ASF DAAC. Alos palsar radiometric terrain corrected low resolution, 2015.

[51] B.K.P. Horn. Hill shading and the reflectance map. *Proceedings of the IEEE*, 69(1):14–47, 1981.

[52] Robert J. Hijmans. *terra: Spatial Data Analysis*, 2023. R package version 1.7-0.

[53] Cesar Aybar, Luis Ysuhuaylas, Jhomira Loja, Karen Gonzales, Fernando Herrera, Lesly Bautista, Roy Yali, Angie Flores, Lissette Diaz, Nicole Cuenca, Wendy Espinoza, Fernando Prudencio, Valeria Llactayo, David Montero, Martin Sudmanns, Dirk Tiede, Gonzalo Mateo-García, and Luis Gómez-Chova. Cloudsen12, a global dataset for semantic understanding of cloud and cloud shadow in sentinel-2. *Scientific Data*, 9(1):782, Dec 2022.

[54] Helen-Nicole Kostis, Kasha Patel, and Dalia B Kirschbaum. New nasa model finds landslide threats in near real-time during heavy rains, 2018. Updated on October 24, 2023.

[55] Sandra Paula Villacorta Chambi, Lionel Fidel Smoll, and Bilberto Luis Zavala Carrión. Mapa de susceptibilidad por movimientos en masa del perú. *Revista de la Asociación Geológica Argentina*, 69(3):393–399, 2012.

[56] Francis Alistair and John Mrziglod. Intelligence for image segmentation (iris). https://github.com/ESA-PhiLab/iris, 2019. Version 1.0.0.

[57] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M. Jorge Cardoso. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. volume 10553, pages 240–248. 2017. arXiv:1707.03237 [cs].

[58] Yang Hong, Robert F. Adler, Andrew Negri, and George J. Huffman. Flood and landslide applications of near real-time satellite rainfall products. *Natural Hazards*, 43(2):285–294, October 2007.

[59] Leijin Long, Feng He, and Hongjiang Liu. The use of remote sensing satellite using deep learning in emergency monitoring of high-level landslides disaster in jinsha river. *The Journal of Supercomputing*, 77(8):8728–8744, Aug 2021.

[60] Wenjie Li, Jingfeng Huang, Lingbo Yang, Yan Chen, Yahua Fang, Hongwei Jin, Han Sun, and Ran Huang. A practical remote sensing monitoring framework for late frost damage in wine grapes using multi-source satellite data. *Remote Sensing*, 13(16), 2021.

[61] Ali Jamali and Masoud Mahdianpari. A cloud-based framework for large-scale monitoring of ocean plastics using multi-spectral satellite imagery and generative adversarial network. *Water*, 13(18), 2021.

[62] Tomáš Řezník, Jan Chytrý, and Kateřina Trojanová. Machine learning-based processing proof-of-concept pipeline for semi-automatic sentinel-2 imagery download, cloudiness filtering, classifications, and updates of open land use/land cover datasets. *ISPRS International Journal of Geo-Information*, 10(2), 2021.

[63] Agencia Peruana de Noticias Andina. Huánuco: deslizamiento por intensas lluvias dañó 200 metros de caminos rurales. 2020.

[64] John Burns Kilbride, Ate Poortinga, Biplov Bhandari, Nyein Soe Thwal, Nguyen Hanh Quyen, Jeff Silverman, Karis Tenneson, David Bell, Matthew Gregory, Robert Kennedy, and David Saah. A near real-time mapping of tropical forest disturbance using sar and semantic segmentation in google earth engine. *Remote Sensing*, 15(21), 2023.

[65] Experiment tracking with weights and biases, 2020. Software available from wandb.com.