

MSc thesis in Geomatics for the Built Environment

Landslide Detection using Random Forest Classifier

Meylin Herrera Herrera

November 2019

LANDSLIDE DETECTION USING RANDOM FOREST CLASSIFIER

A thesis submitted to the Delft University of Technology in partial fulfillment
of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Meylin Herrera Herrera

2019

Meylin Herrera Herrera: *Landslide detection using Random Forest classifier* (2019)

© This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by/4.0/>.

The work in this thesis was made in the:



OTB Group
Department of the OTB
Faculty of Architecture & the Built Environment
Delft University of Technology



Deltares
Independent institute for applied research in the field of
water and subsurface
Delft

Supervisors:	Dr.ir. Mathias Lemmens Dr.ir. Amin Askarinejad
Co-reader:	MSc. Gao Weixiao
External Supervisors:	Dr.ir. Faraz Tehrani Ir. Giorgio Santinelli

ABSTRACT

Landslides are destructive and recurrent natural disasters that cost annually significant social and economic losses all over the world. These events can be induced by natural factors as earthquakes and extreme rainfall, as well as by human intervention, including construction and mining. A primary resource to conduct landslides studies for prediction, risk assessment, and mitigation are historical databases with accurate location of individual events. To increase the location accuracy of those past landslide events, and optimize conventional time- and cost- consuming mapping routines, this study aims to develop an automatic landslide detection method from free-of-charge optical satellite imagery (Sentinel-2) and global Digital Elevation Model (ALOS World3D-30m DEM) using Object-based Image Analysis (OBIA) in combination with Machine Learning (ML).

Existing works have successfully used earth-observation datasets for the generation of landslides databases. Most of them apply **rule-based** techniques using features thresholds that are not global and therefore perform poorly when applied to new regions where the method was not developed. This study presents a first attempt of an automatic method that generalizes to landslides occurring over the entire world without knowledge of their cause or triggering factor.

To obtain a robust method that can deal with the complex characteristics of landslides (e.g. diversity of shapes/sizes, land cover, illumination and spectral variability), we explored OBIA, an image processing technique that has demonstrated better performance than the pixel-based approach, specially when the target objects are bigger than the cell resolution. The developed method consists in cloud-free images acquisition and determination of suitable features for image segmentation and image classification. For the image segmentation, we developed a two-step approach that consists in an initial segmentation using k-means and the **Red/Green Difference** (RGD) as input feature to create homogeneous segments and isolate *landslides* from *non-landslides*. This first approach leads to oversegmentation of *non-landslide* areas and, consequently, to an **imbalanced** dataset. The second step consists in a merging algorithm using Normalized Difference Vegetation Index (NDVI) as input feature to merge homogeneous *non-landslide* segments and balance the dataset.

These two-stages include the setting of parameters as the number of clusters (K) and NDVI thresholds that were experimentally derived. Once the segments are created and the dataset is balanced, a non-parametric supervised classification using Random Forest (RF) was applied to identify *landslide* segments; the main advantage of this classifier is that it can deal with different statistical distributions of features and can handle **imbalanced** datasets. Using a training and testing set of 70% and 30%, our method achieved a precision of 83%, recall of 83%, and f1-score of 83%. We found that topographic features have less influence than spectral ones; however, their exclusion decreases the model performance in about 10%.

Our method is built using entirely open source technologies allowing its applicability and re-usability. For future work, we propose to use our method to detect new landslides and increase the number of training samples. Additionally, we recommend to explore a complementary approach to the merging algorithm to reduce the number of *non-landslide* segments, balance the dataset, and keep accurate classification results while more training images are added to the model.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the people who supported me either directly or indirectly during the execution of my master thesis project.

First, I would like to thank my first supervisor Mathias Lemmens for planning the regular meeting with all my supervisors, his continuous guidance, critical and constructive feedback, and all the shared knowledge as a professor/director of the master program in Geomatics. Many thanks to my second supervisor Amin Askarinejad for its positive and encouraging comments, useful feedback, and valuable discussions we had during my thesis project. I would also like to thank my external supervisors Faraz Tehrani and Giorgio Santinelli, for giving me the opportunity of developing my master thesis at Deltares. Thanks for the regular meetings and the fruitful feedback and discussions we had during my time at Deltares. Thanks as well to my co-reader Gao Weixiao for his additional feedback and helpful comments.

Special gratitude goes to my lovely family. To my husband Lucio, for being part of my life, for our family, and for always be there for me. Thanks for your infinite love, care, and unconditional support. To my angels, Mama and Papa for being my shining stars every day of my life. To my siblings, especially to my sister Merly, who has always been supporting me at all stages of my life. Thanks for your warm encouragement and motivation, your wise words, and your positive energy. To my nephews, especially to Kike for always be present. To my parents-in-law, Lucio and Margarita, for their endless support and motivation in our decisions and adventures. To my husband's cousin, Jorge Collinet, for the valuable discussions and talks that showed me the path towards Machine Learning and Data Science.

Many thanks to all my fellow students for all the memories and experiences we shared in Geomatics. Special thanks to Shenglan, Gabriel, Leiden, Pablo, Ioanna, Giorgos, Melika, Nikos, Vasileios, Teng, Takis, and Qu for the beautiful relation and friendship we built.

Meylin Herrera Herrera

CONTENTS

Glossary	xxi
1 INTRODUCTION	1
1.1 Problem Statement and Motivation	1
1.2 Research Questions	3
1.3 Contribution	3
1.4 Research Scope and Limitations	4
1.5 Methodology Overview	4
1.6 Thesis Outline	5
2 THEORETICAL BACKGROUND AND RELATED WORKS	7
2.1 Image Classification in Remote Sensing	7
2.2 Object-Based Image Analysis	9
2.2.1 Image Segmentation	11
2.2.2 Object-Based Classification	11
2.3 Random Forest Algorithm	12
2.3.1 Early developments	12
2.3.2 Definition	12
2.3.3 Single Decision Trees in Random Forest	13
2.3.4 Main properties of Random Forest	13
2.4 Support Vector Machines	14
2.4.1 Main properties of Support Vector Machines	14
2.5 Imbalanced Learning	15
2.5.1 Handling imbalanced datasets	15
2.5.2 Random Forest for imbalanced datasets	16
2.5.3 Metrics for model performance	16
3 RESEARCH METHODOLOGY	19
3.1 Image Pre-processing	19
3.1.1 Samples set preparation	19
3.1.2 Image set preparation	19
3.1.3 Features computation at pixel level	22
3.2 Image Segmentation	25
3.2.1 Initial Segmentation	25
3.2.2 Merging algorithm	27
3.2.3 Features computation at segment level	30
3.3 Image Classification	31
3.3.1 Training samples and labeling	32
3.3.2 Random Forest Classifier	32
3.3.3 Training phase	34
3.3.4 Model assessment	35
4 DATASETS AND TOOLS	37
4.1 Datasets	37
4.1.1 Sentinel-2 mission	37
4.1.2 Digital Elevation Model	38
4.1.3 Software specifications	38
5 EXPERIMENTS AND RESULTS	41
5.1 Pre-processing	41
5.1.1 Sample set preparation	41
5.1.2 Image set preparation	42
5.2 Image segmentation	44
5.2.1 Initial segmentation	44
5.2.2 Conversion of features per pixel to features per segments	44
5.2.3 Merging algorithm	46

5.3	Image Classification	49
5.3.1	Exploratory analysis	49
5.3.2	Training and testing	49
6	ANALYSIS OF RESULTS	55
6.1	Segmentation Strategies	55
6.1.1	Optimization of the number of segments	56
6.2	Model Training and Classification	56
6.2.1	Final model assessment	57
6.3	validation	58
7	CONCLUSIONS AND FUTURE WORKS	63
7.1	Conclusions	63
7.2	Future Works	64
A	PRE-PROCESSING	77
B	IMAGE SEGMENTATION	79
C	EXPLORATORY ANALYSIS	81

LIST OF FIGURES

Figure 1.1	Methodology Overview. (a) Pre-processing. (b) Image segmentation. (c) Image classification.	6
Figure 2.1	Spatial resolution vs. objects under consideration. (a) Objects size similar to the cell resolution. (b) Objects size bigger than the cell resolution (Modified from Blaschke [2010]).	8
Figure 2.2	Spectral pixel signatures of different objects in Sentinel-2. (a) Agricultural land. (b) Landslide. The picture highlights the limitation of pixel-based approaches when target objects are bigger the cell resolution. Objects are composed of many pixels. Pixel signatures for different objects can be similar. . .	10
Figure 2.3	Single Decision Tree.	13
Figure 2.4	Support vectors and classification margin [OpenCv, 2019]. . .	15
Figure 2.5	ROC curve of three classifiers: A, B and a random classifier (Branco et al. [2016]).	18
Figure 3.1	Cloud-free images pre-processing workflow.	20
Figure 3.2	Image set preparation.	21
Figure 3.3	Gaussian Distribution function (modified from Dougherty [2012]).	29
Figure 3.4	Region growing representation at segment level. The numerical values are hypothetical NDVI values. (1) The algorithm finds the five nearest adjacent neighbors to the seed. (2) The region starts to grow from the seed; the first adjacent segment to be compared is the closer in similarity (closest in NDVI); if the segments satisfy the homogeneity criterion, then the adjacent segment is added to the seed and the NDVI is recalculated using the weighted mean. (3) The algorithm finds the adjacent neighbors of the added segment. (4) Next segment closer in similarity is added to the region, and the NDVI is recalculated. (5), (6), the process is repeated until a stop is reached and no new segment can be added.	29
Figure 3.5	KD-tree data structure (modified from van Oosterom [1999]). . .	31
Figure 5.1	Global coverage of the sample set. The red dots represent the location of the samples images. Given the large-scale of the map, some dots are overlapped in specific areas such as Japan and New Zealand.	42
Figure 5.2	Generation of cloud-free images pre- and post landslide event and heads-up digitization of landslides polygons. (a) Cloud-free images are created using the image composition technique explained in Section 3.1.2 (Landslide L2). (b) Clouds are removed using cloud masking (Landslide L7).	43
Figure 5.3	Comparison of segmentation results using different input features. (a) Segmentation on image difference using RGD (landslide L7); landslide boundaries well defined within an urban area (b) Segmentation on post-event image using NDVI; landslide and building pixels are merged into the same segment. (c) Segmentation on image difference using RGD (landslide L17); landslide boundaries well defined in a remote area. (d) Segmentation on post-event image using bands B4, B4, B2; landslide boundaries (white arrows) are merged to a small deforested area.	45

Figure 5.4	Initial segmentation using two values of k ($L9$). (a) Image difference using RGD. (b) Segmented image, $k=19$; large-scale landslide is correctly segmented. (c) Segmented image, $k=8$, the <i>landslide</i> is undersegmented and merged with <i>non-landslides</i> .	46
Figure 5.5	Pixels to Segments. Features statistic at segment level. (a) Image Difference (b) Segmented image. (c) Post-event image.	46
Figure 5.6	Post-event image of landslides L_0 , L_2 , and L_{16} . The white circle depicts the landslide location.	48
Figure 5.7	Standardized NDVI, VID, and RGD distributions for Landslides L_0 (a), L_2 (b), and L_{16} (c).	48
Figure 5.8	Class distribution and correlations between features: NDVI, Brightness, bands B_4 , and B_2	50
Figure 5.9	Class distribution and correlations between features: Red/Green difference (RGD), $RGD_{deviation}$, Brightness difference (BrightnessD), and $BrightnessD_{deviation}$	51
Figure 5.10	Model performance vs. features. Features represented in Table 5.10 are sequentially input to the model (1-10).	51
Figure 5.11	Confusion matrix (<i>landslides=1, non-landslides=0</i>).	53
Figure 5.12	Final ranking of the landslides diagnostic features using Mean Decrease Impurity of RF.	53
Figure 5.13	Confusion matrix excluding samples with mixed land cover: urban/vegetated (<i>landslides=1, non-landslides=0</i>).	54
Figure 6.1	Image pre-processing and segmentation; sample in a remote area in Italy (L_{17}). (a) Cloud-free pre-landslide image. (b) Cloud-free post-landslide image. (c) Image difference using band ratioing red/green (RGD). (d) Image segmentation.	55
Figure 6.2	NDVI histograms for <i>landslide</i> and <i>non-landslide</i> segments.	57
Figure 6.3	Training samples with mixed land cover: urban-vegetated.(a) Landslide L_2 . (b) Landslide L_7	58
Figure 6.4	Model validation of landslides in New Zealand (event date: 2016-11-15). (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. Yellow circle, landslides detected. Red circle, landslides not detected.	59
Figure 6.5	Model validation on area with old landslides. (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. No landslides detected.	59
Figure 6.6	Model validation in the Netherlands. (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. Red arrows indicate segments with similar spectral information than landslides. No landslides detected	60
Figure 6.7	Model validation in Nepal. (a) Landslides detection using the NASA tool. (b) Image difference using RGD. (c) Post-event image. White areas, high change between pre- and post image. Dark grey areas, low change. Yellow circle, landslides detected for our tool. Red circle, landslides not detected.	61
Figure A.1	Landslides database. An example of the landslides tables stored in PostgreSQL. The original data includes additional information such as the length and width of the major landslide recognized in the image, the landslide category (e.g. landslide, mudslide), landslide setting (e.g. natural slope, above river), location name (e.g. Mocoa, Colombia; Java, Indonesia), water body proximity (e.g. close to river, lake or coast).	78

Figure B.1	Image set preparation.	79
Figure C.1	Class distribution and correlation between features: Slope_mean, Slope_max,and Relative_relief.	81

LIST OF TABLES

Table 2.1	OBIA vs. pixel-based approach.	10
Table 2.2	Confusion matrix for binary classification. Tp are the true positives predictions, Tn are the true negatives, Fp are the false positives, and Fn are the false negatives	16
Table 3.1	Initial landslide diagnostic features.	22
Table 4.1	Sentinel-2 spectral and QA bands.	37
Table 4.2	Sentinel-2 mission details.	38
Table 4.3	DEM ALOS description.	38
Table 5.1	Landslide sizes.	41
Table 5.2	Landslide database example. The location accuracy “bbox center” means that the geographical coordinates represent the center of the bounding box used to download the image (see Figure 3.2). For an extended version of the database see Appendix A	42
Table 5.3	Quality assessment matrix for the samples set. Gl is the geographical location of the major landslide recognized in the image, C is the cloud percentage, and S is the snow percentage.	43
Table 5.4	Summary of the image and sample sets inventory.	43
Table 5.5	Initial segmentation parameters.	44
Table 5.6	Attribute table example ($L17$).	47
Table 5.7	Landslide diagnostic features at segment level.	47
Table 5.8	Merging algorithm using fixed NDVI thresholds.	48
Table 5.9	Merging algorithm using the distributions of spectral features. A sequential merging criteria is added: NDVI, VID, RGD, and a fine-grained threshold.	49
Table 5.10	Initial ranking of landslides diagnostic features.	50
Table 5.11	Optimal RF hyperparameters.	52
Table 5.12	Final model performance.	52
Table 5.13	Model performance excluding samples with mixed land cover: urban/vegetated.	54

List of Algorithms

3.1	Landslide Image Pre-processing (GEE)	21
3.2	Similarity Measure	28
3.3	Region Growing	30

ACRONYMS

ALOS	Advanced Land Observing Satellite	38
ANN	Artificial Neural Networks	7
AUC	Area Under the Curve	18
AW _{3D30}	ALOS World 3D-30m	4
BRF	Balanced Random Forest	16
BrightnessD	Brightness Difference	24
B ₂	Blue band	38
B ₄	Red band	44
B ₃	Green band	44
B ₁₀	high atmospheric absorption band	38
CART	Classification and Regression Tree	32
CNN	Convolutional Neural Networks	8
DEM	Digital Elevation Model	2
ESA	European Space Agency	37
DSM	Digital Surface Model	38
GEE	Google Earth Engine	2
GIS	Geographical Information System	1
GNDVI	Green Normalized Difference Vegetation Index	23
GSD	Ground Sample Distance	2
ID ₃	Iterative Dichotomiser	32
JAXA	Japan Aerospace Exploration Agency	38
KD-tree	K-Dimensional tree	30
KNN	K-Nearest Neighbor	7
NDVI	Normalized Difference Vegetation Index	9
MDA	Mean Decrease Accuracy	14
MDI	Mean Decrease Impurity	14
ML	Machine Learning	1
MLC	Maximum Likelihood Classification	7
NIR	Near-Infrared	23
OBIA	Object-Based Image Analysis	2
OOB	Out-of-Bag	34
PRISM	Panchromatic Remote-sensing Instrument for Stereo Mapping	38
GEOBIA	Python-Based Open Source System for Geographic Object-Based Image Analysis	39
QA ₆₀	Quality Assessment	22
R/G	Red/Green ratio	23
RGD	Red/Green Difference	24
RF	Random Forest	4
RIOS	The Raster Input and Output Simplification	39

ROC Receiver Operating Characteristic	18
SWIR Short-Wave Infrared	37
SVM Support Vector Machines	5
TOA Top-Of-Atmosphere	37
VID Vegetation Index Difference	24
WRF Weighted Random Forest	16

GLOSSARY

Bootstrap Aggregation is a method that uses multiple versions of a predictor and using them to get an aggregated predictor. It consists in making a random selection of samples with replacement (or re-substitution) in such a way that the record already selected for training is placed back into the original dataset and can be used again to build a new training set. 12, 34

change features are those features derived from applying the image difference techniques and that are related with changes between the pre- and post-landslide image (e.g. Red/Green difference (RGD), Vegetation Index difference (VID), Brightness difference (BrightnessD)). 47, 58

coarse-grained threshold is a flexible threshold. For instance, the similarity threshold between two segments is coarse-grained if it is ≥ 0.1 . 47, 48

contextual or relative to the information contained in the image. 2, 11, 27, 31, 49, 58, 63

ensemble is a Machine Learning technique that combines multiple learning algorithms to generate a model with better predictive performance than the single components of the model. Examples are Random Forest and Boosting algorithms. 7, 12, 13, 16, 32, 34

fine-grained threshold is a strict or non-flexible threshold. The similarity threshold between two segments is fined-grained if it approximates to 0 (e.g. $0 < \text{threshold} < 0.1$). 28, 47, 48, 56

geomorphological settings are landforms derived from the physical evolution of the landscape. Examples of geomorphological settings are river valleys, fluvial terraces, alluvial fans, deltas, hilly topography, and landforms derived from eolian, coastal-marine and glacial processes. 1, 3, 26, 57, 64

heads-up digitization is the manual digitization by tracing a mouse over features displayed on a computer screen ¹. 19

imbalanced dataset, for a binary classification problem, a dataset is said to be imbalanced when one of the classes constitutes the majority objects, or it considerably outnumbers the other class.. v, 5

knowledge-based features are features commonly used by experts during visual image interpretation to identify landslides. 22

landslides diagnostic features are knowledge-based features that give key information to characterize and identify landslides. Examples are the Normalized Difference Vegetation Index (NDVI), Brightness, and Green Normalized Difference Vegetation Index (GNDVI), slope angle, terrain curvature, slope direction. 2, 4, 5, 20, 22, 29, 30, 35, 45, 49, 52, 57

non-local or global feature. Feature mean calculated using all pixels contained in the image. 28, 47

overfitted model refers to a model which does not generalize well to samples not encountered during the training. 14

¹ <https://support.esri.com/en/other-resources/gis-dictionary/term/0c14e614-30c2-4bee-a1e4-4ff2c60d4626>

Red/Green Difference It is a landslide diagnostic feature calculated as the difference between the Red/Green ratio in t_1 (pre-event) and the Red/Green ratio in t_2 (post-event) to produce a further image which represents the change between the two times. [v](#)

rule-based classification technique that consists in manually define decision trees using features thresholds. [v](#), [3](#), [60](#)

spectral features features derived from optical satellite images (e.g. NDVI, brightness). [9](#), [22](#), [25](#), [28](#), [47](#), [56](#), [59](#), [60](#), [63](#)

topographic features morphometric features derived from Digital Elevation Models (e.g. slope, relative relief). [22](#)

variance in Machine Learning, refers to the variability in the predictive performance of a Machine Learning model when it is tested on the training and the unseen data. A model with high variance performs well on the training data but can not generalizes on the test or unseen data. [12](#), [13](#), [33](#), [34](#), [52](#)

Natural disasters are destructive forces that cost annually significant social and economic losses all over the world. In particular, landslides are one of the most recurrent and widespread natural disasters as they are induced not only by natural factors as earthquakes and extreme rainfall but also by human intervention, including construction and mining. According to [Froude and Petley \[2018\]](#), from 2004 to 2016, about 56,000 human lives were lost in 4,862 rainfall-induced landslide events distributed worldwide, with Asia representing the dominant geographical area.

Landslides can occur in a wide diversity of [geomorphological settings](#) from coastal regions and river valleys to hilly areas. They are not restricted to a particular type of lithology and can be associated with different types of land cover, including urban areas, forests, bare-lands, grasslands, and agricultural lands. The large variety of landslides and their dependency on natural factors limit the reliability of their prediction, risk assessment, and mitigation. Nevertheless, with the advent of the world wide web, the recent advances in Geographical Information System (GIS) technologies, the availability of public earth-observation data, and the fast growth of Machine Learning (ML) during the last decades, many efforts have been made to reduce uncertainties in landslide-related studies and minimize mortalities and economic losses induced by landslides [[Pradhan and Lee, 2010](#); [Nithya and Prasanna, 2010](#); [Subhashini and Premaratne, 2013](#); [Zhou et al., 2018](#); [Tehrani et al., 2019](#)].

A primary resource to conduct landslide studies at large, regional, and local scales, is a landslide inventory map [[Raspini et al., 2016](#)]. [Van Westen \[2016\]](#) defines a landslide inventory as a collection of landslide features in a particular area for a certain period, with spatial information related to the location (as points or polygons), preferably combined with attribute information such as date of occurrence, triggering factor, and landslide size.

1.1 PROBLEM STATEMENT AND MOTIVATION

Creating landslides inventories maps is the first step for the training and testing of landslide forecasting studies [[Danneels et al., 2007](#)]. [Van Westen et al. \[2006\]](#) state that the largest source of error in landslide susceptibility and risk maps for land-use planning and other mitigation measures is the limited availability of existing landslides inventories. The main challenge of generating such landslide databases is the estimation of the accurate spatial location of individual landslide events and their temporal occurrences. Global landslides catalogs as the one developed by NASA Goddard Space Flight Center (GSFC) [[Kirschbaum et al., 2010](#)], provide initial insights into the spatiotemporal statistical trends in the worldwide landslides distribution. [Kirschbaum et al. \[2010\]](#) report significant uncertainties when assigning geographic coordinates to a landslide event. To deal with these uncertainties, the authors set a radius of confidence (which spans from tens of meters to tens of kilometers) to the location, indicating the potential area over which the landslide may have occurred [[Herrera, 2018](#)]. The accurate location of landslides events is critical for the prediction of landslides. Controlling factors such as soil composition, soil moisture, vegetation index, and topographic features (e.g. slope, aspect) are strongly dependent on the geographic location of the landslide as their values can drastically vary within a few meters.

Satellite remote sensing technology plays a central role in the decision-making process for disaster management. Implementing warnings and mitigation strategies are often challenging if developed without having robust ground observation data. Therefore, earth-observation data has a higher value when ground data is limited or unavailable [Kansakar and Hossain, 2016]. Remote sensing has proved its effectiveness in generating landslide inventories [Martha et al., 2011], assisting traditional time-consuming mapping methods that mostly rely on field survey and visual interpretation of aerial photographs [Lu et al., 2011]. Multi-temporal imagery for natural disaster assessments has become an alternative source to deal with areas with limited access to public geo-data. The use of platforms such as Google Earth Engine (GEE) ¹, USGS Earth Explorer ² and Copernicus ³ allows public access to satellite-based datasets, which make possible time series analysis on a global scale. Imagery collections from the Sentinel program, provide a continuous record of satellite-based observations, and together with Landsat imagery, they are primary sources of medium-spatial resolution datasets for monitoring global change. Sentinel-2, with Ground Sample Distance (GSD) of 10m, has a broad list of applications, including detecting land cover change, estimating chlorophyll concentrations, monitoring geohazards, and providing emergency response in natural disaster management [Lemmens, 2015].

The value of earth-observation data for landslides detection has continuously increased during the last two decades. The global geographical coverage, high temporal frequency, and low cost, mixed with a broad selection of spatial and spectral options, highlight the use of remotely sensed imagery for change detection in different applications [Chen et al., 2012]. Hölbling et al. [2015] explain that satellite data is crucial for detecting landslides after triggering events, mainly when they occur in remote or hardly accessible areas.

To fully exploit the wide range of existing optical imagery, remote sensing has evolved into a multidisciplinary field where Machine Learning (ML) has caused a significant impact [Camps-Valls, 2009]. Some of the meaningful advances in developing ML-based methodologies for remote sensing applications are image classification and automatic extraction of information by computational and statistical methods [Lary et al., 2016]. Image detection methods in remote sensing include classical methods focused on pixel-based approaches [Cheng et al., 2004; Danneels et al., 2007; Tsangaratos and Ilia, 2014] and advances techniques such as the Object-Based Image Analysis (OBIA), which has proved high performance in many applications [Platt and Rapoza, 2008; Blaschke, 2010; Martha et al., 2011; Lu et al., 2011; Martha et al., 2012; Feizizadeh et al., 2017], especially with medium-high and very-high-resolution satellite imagery. In contrast to pixel-based approaches, OBIA allows the integration of several **landslides diagnostic features** such as spectral, spatial, and **contextual** features [Martha et al., 2010] that better resemble visual interpretation of aerial photographs [Feizizadeh et al., 2017] and reduce the influence of the single-pixel reflectance [Blaschke, 2013].

Although conventional methods to create landslide inventories have been used with standardized procedures for several decades, they are resource-intensive and time-consuming [Raspini et al., 2016]. To optimize these time- and cost- consuming mapping routines exploiting the use of free-of-charge earth-observation data, this study aims to develop an automatic landslide detection method from Sentinel-2 and Digital Elevation Model (DEM) using Object-Based Image Analysis OBIA in combination with Machine Learning (ML).

¹ <https://earthengine.google.com/>

² <https://earthexplorer.usgs.gov/>

³ <https://www.copernicus.eu/en>

1.2 RESEARCH QUESTIONS

Based on the research objective and the problem introduced in the previous section, the main research question of this thesis is formulated:

How to detect landslides using Machine Learning?

To answer the main research question, the following sub-questions are addressed:

- To what extent can landslides be detected using Sentinel-2 in combination with a worldwide Digital Elevation Model?
- What are the relevant landslide diagnostic features?
- What is the best segmentation strategy?
- How to exploit features per pixels to produce features per segments?
- What is the most appropriate Machine Learning technique?
- What is the accuracy of the most appropriate Machine Learning technique?

1.3 CONTRIBUTION

The main contributions of this study are:

- Model generalization. Only few works have combined OBIA and ML for landslide detection using remote sensing data [Danneels et al., 2007; Stumpf and Kerle, 2011; Parker, 2013]; most of them apply rule-based techniques using features thresholds (e.g. Vegetation index) [Lu et al., 2004; Blaschke, 2010; Martha et al., 2012; Keyport et al., 2018]. Such models use training data and thresholds that are region-specific and consequently perform poorly when applied to new regions where the method was not developed. In contrast to previous research, our method is trained with satellite images from different areas, landscapes, and geomorphological settings and with different types of landslides not tied to a specific triggering factor (e.g. earthquakes, rainfall). This variety of training samples allowed the generation of a model that generalizes to landslides occurring worldwide without knowledge of the cause or triggering factor.
- Multi-scale approach. One of the main difficulties in the application of any image segmentation technique is to deal with different object sizes. Some works use single segmentation scales [Espindola et al., 2006; Yu et al., 2006; Zhou and Troy, 2009], while others investigate different approaches to select the optimal input parameters for multi-scale segmentation algorithms [Benz et al., 2004; Johnson and Xie, 2011; Martha et al., 2011] that are implemented within commercial-software (e.g. eCognition Definiens Developer ⁴, Envi Zoom, and ERDAS Objective). Using open-resource alternatives, we use a two-step segmentation approach combining k-means clustering for initial and detailed segmentation, and region growing algorithm at segment level to merge homogeneous segments. This strategy allowed to isolate landslides of different sizes (landslides not merged with non-landslide segments) and balance the training set;

⁴ <http://www.ecognition.com/>

- Free-of-charge resources. The re-usability and applicability of previous landslides detection methods are restricted due to the use of commercial software (e.g. eCognition) and datasets (e.g. Quickbird, Spot, IKONOS, WorldView-2). Our method is built using free-of-charge software (Python and GEE) and datasets (Sentinel-2, ALOS World 3D-30m (AW3D30) DEM).
- Semi-automated implementation. A semi-automated tool that includes the generation of cloud-free images, the extraction/computation of features, and the segmentation and classification of optical satellite images is developed to test our proposed method. This implementation can be used to assist landslides experts/non-experts in detecting new landslides events and improve existing databases at low costs.

1.4 RESEARCH SCOPE AND LIMITATIONS

The following considerations are made to explain the research scope and highlight the limitations of this study:

- The main product of this study is a method to detect landslides from optical satellite imagery and DEM using OBIA in combination with Random Forest (RF). The semi-automatic tool developed to test and validate our method consists of a pipeline that includes the image pre-processing stage, image segmentation, and image classification. Since the pre-processing stage was implemented using Google Earth Engine (GEE), it could only serve for research, development, or educational purposes unless the user signs other terms and agreements with Google. The prototype aimed to validate the methodology and not at being a production-ready software;
- Due to the GSD of the available dataset (Sentinel-2), this study focuses on large (100px) to catastrophic-scale landslides with minimum size of approximately 10,000 m² (100m x 100m);
- Due to the temporal coverage of Sentinel-2 imagery, the model is exclusively trained with landslides occurred from June 2015 and onwards;
- The developed method has limitations in regions with perennial snow, high sedimentation rate, densely built-up areas, and regions characterized by sparse or non-vegetation;
- This study aims at detecting landslides, not their delineation;
- This study does not aim to compare results from different image detection techniques (pixel-based vs. OBIA) or ML algorithms;

1.5 METHODOLOGY OVERVIEW

Based on the research questions (Section 1.2) and the principles of OBIA, the following methodology workflow is defined:

1. Pre-processing. This stage involves the sample set preparation, generation of cloud-free images, the extraction of spectral and topographic features from Sentinel-2 and AW3D30 DEM, and the computation of landslides diagnostic features at pixel level (Figure 1.1 a).
2. Image segmentation. It is the first step towards the application of OBIA. Blaschke [2010] defines image segmentation as the subdivision of an image into spatially continuous, disjoint, and relative homogeneous regions that refer to

segments. This stage is implemented as a two-step approach: (a) an initial segmentation using a k-means implementation; (b) a merging algorithm using a region-growing implementation (Figure 1.1 b).

3. Image classification. Once segments with features statistics are obtained, the image is classified by assigning each segment to a class. The classification problem of this study is binary with the classes *landslide* and *non-landslide*.

The image classification is conducted using supervised ML, specifically RF algorithm. It involves the training samples preparation and labeling, the exploratory analysis to understand the statistical distribution of the data, the model training, and testing, and finally, the assessment of the model performance. The adopted approach for classification is conducted by iteratively adding features and assessing the model performance at each step (Figure 1.1 c)

1.6 THESIS OUTLINE

This thesis document is structured as follows:

- Chapter 2 reviews the related works and theoretical background of two main image detection techniques in remote sensing. The chapter also describes the OBIA workflow and general concepts related to Random Forest (RF), Support Vector Machines (SVM) and imbalanced datasets;
- Chapter 3 addresses the mathematical foundations, concepts, and algorithms implemented for the proposed method including the pre-processing techniques for image analysis, computation of landslides diagnostic features, image segmentation, and image classification;
- Chapter 4 describes the dataset and tools;
- Chapter 5 describes the implementation details of the methodology, the experiments conducted, and presents the results;
- Chapter 6 addresses the analysis of the results;
- Chapter 7 presents a summary of the conclusions by answering the main research questions. The chapter and the thesis document end with recommendations for future works.

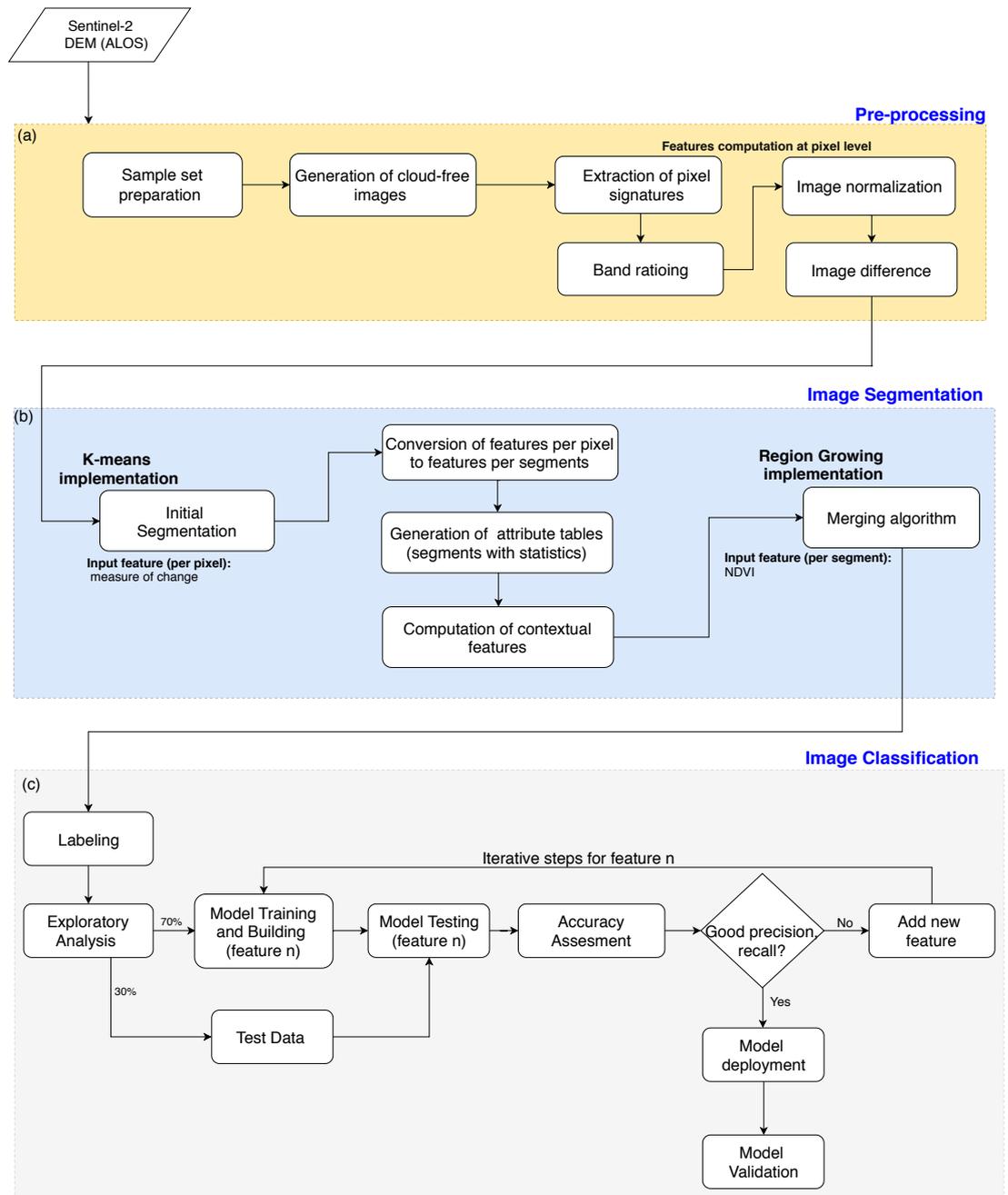


Figure 1.1: Methodology Overview. (a) Pre-processing. (b) Image segmentation. (c) Image classification.

2

THEORETICAL BACKGROUND AND RELATED WORKS

This chapter aims to give an overview of the relevant theory and previous scientific works related to this study. [Section 2.1](#), discusses the developments of two main image detection techniques for remote sensing applications (pixel-based vs. [OBIA](#)). [Section 2.2](#) addresses the concept of [OBIA](#), its workflow, and its relevance for landslide detection. [Section 2.3](#) reviews the definition, early developments, and main properties the Random Forest ([RF](#)) algorithm. [Section 2.4](#) gives a general introduction to [SVM](#), a [ML](#) technique that similar to [RF](#), is commonly used for object-based classification in remote sensing. Finally, an introduction to imbalanced learning and the suitable metrics for evaluating model performance are presented in [Section 2.5](#).

2.1 IMAGE CLASSIFICATION IN REMOTE SENSING

Statistical pattern recognition and the use of image classification techniques in remote sensing have been widely researched over several years [[Fu, 1983](#); [Lemmens, 1994](#); [Chen and Ho, 2008](#); [Dougherty, 2012](#); [Khorram et al., 2016](#)]. First efforts were mostly dedicated to multispectral classification using pixel-based approaches where individual pixels are handled as separate objects without exploring their spatial relationships with adjacent pixels [[Lemmens, 2011](#)]. Statistical pattern recognition deals with accurately classifying a pattern into one of several classes [[Chen and Ho, 2008](#)]. In a pixel-based classification problem, each pixel is assumed to be an individual object and to belong to a single class.

Data-driven techniques are of primary relevance for image processing in remote sensing. In particular, supervised and unsupervised learning are the two main approaches for multispectral classification. While in unsupervised classification, the pixels are grouped into clusters based on their spectral properties before the class assignment, in supervised classification the spectral signatures of each pixel are first linked to classes during a training phase. Supervised classification works on the principle of using labeled information about class membership of single pixels to create a model that can generalize to the whole image or a set of images [[Camps-Valls, 2009](#)]. The *prior* knowledge is obtained by collecting ground data for a limited number of areas that represent the training samples [[Lemmens, 2011](#)].

In order to compare the pixel values to those of the training samples, and allocate individual objects (pixels in pixel-based approach) to the most likely class [[Lemmens, 1994](#)], different classification algorithms have been used for a great variety of applications including land-use and land cover classification, object detection, aquatic vegetation classification, water quality mapping, crops monitoring, urban planning, and disaster management. Some of the most commonly applied methods are statistical learning algorithms such as Maximum Likelihood Classification ([MLC](#)) [[Strahler, 1980](#); [Rajan et al., 2008](#); [Khorram et al., 2016](#)], K-Nearest Neighbor ([KNN](#)) [[Cover et al., 1967](#); [Grabowski et al., 2003](#)], Minimum Distance to Mean Classifier, and Parallelepiped Classifier [[Khorram et al., 2016](#)], as well as more advanced [ML](#) algorithms such as [ensemble](#) methods or techniques that combine multiple learning algorithms (e.g. [RF](#), Boosting) [[Briem et al., 2002](#); [Ham et al., 2005](#)], [SVM](#) [[Camps-Valls et al., 2006](#); [Inglada, 2007](#); [Camps-Valls et al., 2008](#)], and Artificial Neural Networks ([ANN](#)) [[Del Frate et al., 2007](#); [Jensen et al., 2009](#)]. Lately, Deep Learning models start being used in a variety of image classification problems [[Krizhevsky et al.,](#)

2012; Chan et al., 2015; Hu et al., 2015]. For the specific case of satellite imagery classification, Nguyen et al. [2013] defined an architecture of Convolutional Neural Networks (CNN) for image classification that does not require any pre-processing or feature extraction process. Maggiori et al. [2016] proposed a fully CNN for dense pixel-wise classification of satellite imagery to produce fine-grained classification maps. Kersbergen [2018] used CNN for automated building damage classification from radar and very-high-resolution optical imagery.

In the context of land cover classification and landslide detection, a variety of pixel-based studies have been conducted using different classification techniques and earth-observation datasets. Danneels et al. [2007] generated an automated procedure to detect landslides from multispectral remote sensing images using MLC. Similarly, Sekertekin et al. [2017] used MLC to derive land-use and land cover images from Landsat 8 and Sentinel-2. Tsangaratos and Ilia [2014] developed a supervised ML spatial tool to detect terrain deformation induced by landslides occurrences using Google Earth Engine (GEE). Enderle and Weih Jr [2005], integrating supervised and unsupervised classification methods, generated an accurate land cover classification map using Landsat 7 imagery. Lee and Lee [2006] implemented a change detection technique to detect landslide using KOMPSAT-1 satellite imagery.

With the continuous improvement of the spatial resolution of remote sensors (e.g. IKONOS, SPOT-5, Quickbird, WorldView), new needs for efficiently extracting information from high-resolution and very-high-resolution satellite images emerged. For applications as low-resolution land cover classification where the target objects approximate the size of the cell resolution (Figure 2.1a), it is assumed that pixels belong to the same land cover class if they are close in the spectral feature space [Blaschke, 2013]; however, in the presence of higher resolution imagery (Figure 2.1b), the same assumption fails, especially in more complex environments composed of multi-scale objects. The two situations illustrated by Blaschke [2010] in Figure 2.1, require different approaches to efficiently extract the image information. In case (a), for medium-low resolution imagery, a pixel by pixel technique might be suitable, while in case (b), for high-resolution imagery, the need for a new approach that makes use of spatial concepts and considers the regionalization of neighboring pixels into groups of pixels is required. As a result, OBIA technique is developed to address the pixel-based limitations.

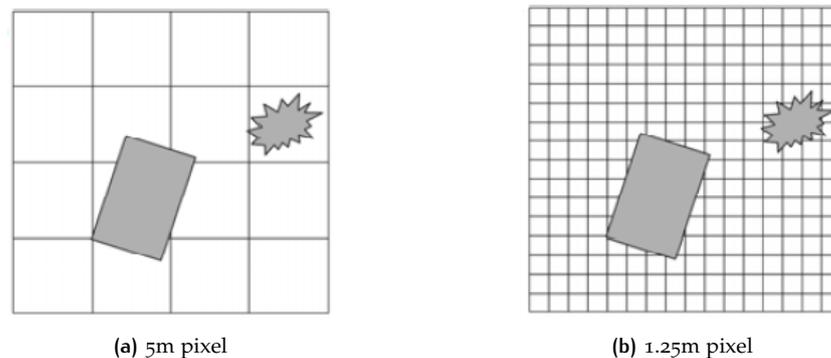


Figure 2.1: Spatial resolution vs. objects under consideration. (a) Objects size similar to the cell resolution. (b) Objects size bigger than the cell resolution (Modified from Blaschke [2010]).

One of the first works in the field of object-based analysis was published by Blaschke et al. [2000]. Their study presented the OBIA method in the context of GIS and remote sensing, and explained the perspectives for environmental applications. Blaschke et al. [2004] presented a new contextual approach using image segmentation for object-based classification. Some years later, a solid overview of the development of object-based methods and their application in remote sensing and GIS is presented by Blaschke [2010]. The author clearly states the limitations

of the pixel paradigm and the considerable progress that **OBIA** methods were making towards a spatial information extraction workflow, as it is required for spatial planning and in many monitoring programs.

Several works have been dedicated to proving the advantages of **OBIA** over the pixel-based approach. [Behling et al. \[2014\]](#) proposed a method that uses pixel-based thresholds and **OBIA** to investigate temporal changes in Normalized Difference Vegetation Index (**NDVI**) and identify landslide-related land cover changes based on RapidEye satellite imagery. [Platt and Rapoza \[2008\]](#) compared results from **MLC** with results from **OBIA** for a mixed urban-suburban-agricultural landscape. [Key-port et al. \[2018\]](#) conducted a comparative analysis of pixel-based and object-based detection of landslides from very-high-resolution images. [Weih and Riggan \[2010\]](#) compared object-based classification with supervised and unsupervised pixel-based classification using SPOT-5 satellite imagery and a high-resolution color infrared digital orthophoto.

2.2 OBJECT-BASED IMAGE ANALYSIS

OBIA aims at grouping neighboring pixels in regions before conducting the classification. In the context of remote sensing applications, this technique provides adequate and automated methods for the analysis of high-resolution and very-high-resolution satellite imagery by describing the image reality using the object diagnostic characteristics [[Lang, 2008](#)]. The main advantage of **OBIA** is the integration of spectral, textural, and spatial/contextual features [[Martha et al., 2010](#); [Lu et al., 2011](#)] to group image pixels into homogeneous and meaningful objects [[Lu et al., 2011](#)]. In contrast to image-pixels, an image-object can be assigned to valid corresponding real-world objects [[Lang, 2008](#)].

The applicability of **OBIA** in the context of landslides detection has been proved in previous works. [Martha et al. \[2012\]](#) conducted an object-based analysis of multi-temporal panchromatic images for the creation of historical landslide inventories. [Blaschke et al. \[2014\]](#) established a semi-automated **OBIA** methodology for locating landslides in north-western Iran. [Hölbling et al. \[2015\]](#) developed a semi-automatic **OBIA** approach for landslide detection in northern Taiwan based on high-resolution satellite data and **DEM**. [Feizizadeh et al. \[2017\]](#) presented an **OBIA** methodology for landslide-related change detection from multi-temporal satellite images using spatial and **spectral features**, and applying fuzzy logic membership functionality.

Due to the out-performance of **OBIA** over the pixel-based approach [[Blaschke, 2010](#); [Weih and Riggan, 2010](#); [Blaschke, 2013](#); [Lu et al., 2011](#); [Hölbling et al., 2015](#)] for classification of complex geospatial objects (different size/shapes, high spectral and illumination variability), the former state-of-the-art image detection technique was selected to develop this study. Landslides have complex shapes, and spectral properties very close to those of other natural objects such as river sand deposition and rocks outcrop [[Martha et al., 2010](#)], or human-intervened objects as agricultural lands and deforested areas ([Figure 2.2](#)).

[Table 2.1](#) summarizes the main differences between **OBIA** and pixel-based approaches. The table also includes the advantages of **OBIA** over pixel-based when the target objects are bigger than the cell resolution ([Figure 2.1b](#) and [Figure 2.2](#)).

Although the input datasets of this study have a medium-spatial resolution (**GSD** = 10m), **OBIA** is still a suitable approach as the target object sizes are considerably larger than the cell resolution. For a homogeneous object to be identified, a minimum object size of approximately $100px$ is suggested. Therefore, considering that the **GSD** of Sentinel-2 is 10m, the minimum landslide size considered for this research is approximately $10,000m^2$ (see [Section 1.4](#)).

The two basic principles of **OBIA** involve segmentation and classification. [Martha et al. \[2010\]](#) state that **OBIA** is effectively a combination of segmentation to derive

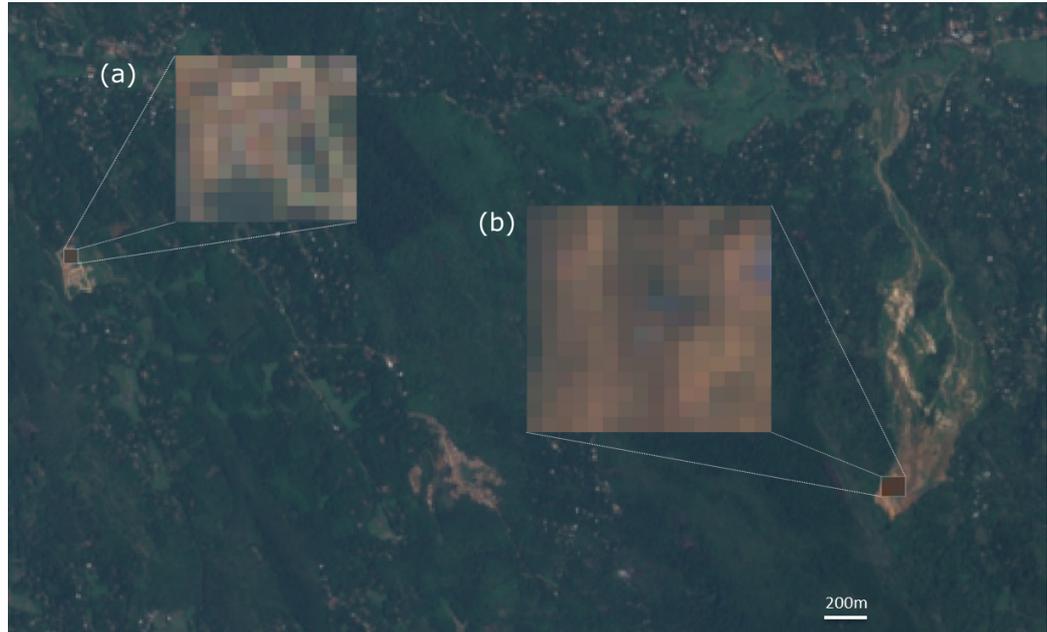


Figure 2.2: Spectral pixel signatures of different objects in Sentinel-2. (a) Agricultural land. (b) Landslide. The picture highlights the limitation of pixel-based approaches when target objects are bigger the cell resolution. Objects are composed of many pixels. Pixel signatures for different objects can be similar.

Characteristics	Pixel-Based	OBIA
Object	Pixels values	Segments statistics
Class	Each pixel belongs to a class	Each segment belongs to a class.
Advantages of OBIA.	Overcomes the mixed pixel problem within the same class Overcomes the so-called salt-and-pepper effect Easily integrated within GIS Reduces computational classifiers loads Can address multiple scales Overcomes the problem of similar pixels for different classes	

Table 2.1: OBIA vs. pixel-based approach.

image primitives, and their subsequent classification based on features calculated from the extracted objects.

2.2.1 Image Segmentation

Image segmentation defines the building blocks of OBIA [Lu et al., 2004] as the accuracy of the classification process is directly influenced by the segmentation results [Lemmens, 2011; Feizizadeh et al., 2017]. According to Blaschke [2010], the homogeneity of the pixels can be determined by one or more spectral criteria in one or more dimensions of a feature space; that means that when compared with single pixels, the segments will have additional spectral information including for instance, mean, median, standard deviation, minimum, and maximum values per band. Besides the diversification of spectral values, a more significant advantage of segmentation is the inclusion of additional spatial and contextual information for target objects, including the neighborhood relationship [Blaschke, 2010; Benz et al., 2004; Hay and Castilla, 2006].

Image Segmentation techniques were developed in the 1980s [Haralick and Shapiro, 1985], deriving in numerous segmentation algorithms in the context of image processing and computer vision [Pal and Pal, 1993]. For the specific case of segmentation of remote sensing data, Li et al. [2014] explain that spatial-contextual information has been added to the original segmentation algorithms. Among the most common techniques highlight region-based segmentation algorithms [Mannan and Ray, 2003], Markovian methods [Jackson and Landgrebe, 2002], Watershed methods [Salembier Clairon et al., 1998], hierarchical algorithms [Dalla Mura et al., 2010], and k-means clustering implementations [Wang et al., 2010; Clewley et al., 2014; Shepherd et al., 2019].

One of the main challenges for segmentation approaches is to segment objects from different sizes in the same image. Blaschke et al. [2000] state that segments in an image do not represent meaningful objects at all scales for any application. To address the multi-scale issue, Baatz [2000] developed a multi-resolution segmentation approach based on a region merging technique implemented in the commercial software eCognition; the approach works as a local optimization procedure that simulates the synchronous growth of segments over a scene. According to Blaschke [2010], most of the works referred to as OBIA originated around eCognition in the early's 2000.

2.2.2 Object-Based Classification

Object-based classification is defined by Castilla and Hay [2008] as the process of allocating image-object (segments) to geo-object classes based on both the internal features of the objects and their mutual relationships. According to Martha et al. [2011], the use of thresholds is a primary requirement for image classification using earth-observation data. The most common object-based classification approach employed in previous works is the application of rule-based techniques [Baraldi et al., 2006; Martha et al., 2012; Moosavi et al., 2014; Keyport et al., 2018] that can involve different segmentation, classification, shape adaptation, and merging steps [Martha et al., 2011]. Danneels et al. [2007] explain the rule-based classification as a step-wise elimination with several decision rules based on knowledge of the target object properties.

More advanced techniques for object-based classification have been investigated during the last decade. ML classifiers such as RF and SVM are among the most commonly applied algorithms for object-based land cover classification [Peña et al., 2014; Qian et al., 2015; Li et al., 2016]. A mixed RF and OBIA classification scheme was developed by Lebourgeois et al. [2017] to combine multi-source datasets (Sentinel-2 time series, very-high-resolution satellite imagery, and DEM) and generate high-

resolution land-use classification. Using very-high-resolution satellite imagery from four case studies, [Stumpf and Kerle \[2011\]](#) explored the applicability and performance of [RF](#) in combination with object-oriented analysis for image classification in the context of landslides detection. [Tzotsos and Argialas \[2008\]](#) evaluated the effectiveness of [SVM](#) approach for multi-class classification based on primitive image objects produced by a multi-resolution segmentation algorithm. [Ma et al. \[2017\]](#) conducted a literature review based on 173 scientific publications in object-based classification for land cover mapping; they concluded that [RF](#) exhibits the best performance [[Stumpf and Kerle, 2011](#); [Duro et al., 2012](#)] followed by [SVM](#) [[Duro et al., 2012](#)]. [Parker \[2013\]](#) conducted an object-based segmentation and made a comparative analysis of the performance of [RF](#) and [SVM](#) for landslide detection using Worldview-2 imagery. The authors demonstrated that [RF](#) classification achieved better results in terms of model performance and processing cost.

2.3 RANDOM FOREST ALGORITHM

2.3.1 Early developments

The Random Forest ([RF](#)) algorithm was for the first time introduced by [Ho \[1995\]](#) and it is nowadays one of the most robust and widely used algorithms within a wide range of fields [[Pretorius et al., 2016](#); [Biau and Scornet, 2016](#)] including remote sensing [[Stumpf and Kerle, 2011](#); [Peña et al., 2014](#); [Ma et al., 2017](#); [Lebourgeois et al., 2017](#)]. Inspired by the limitations of single decision trees, and following the principles of the stochastic modeling [[Kleinberg, 1990](#); [Berlind, 1995](#)], [Ho \[1995\]](#) proposed a method to construct tree-based classifiers whose capacity can be arbitrarily expanded to increase both training and testing set accuracy. [Amit and Geman \[1997\]](#) presented a new approach for shape recognition that allowed to generalize well to samples not seen during the training; the method was based on growing binary classification trees using a large number of randomly selected geometric features to find the best split at each node of the tree. Influenced by the works of [Ho \[1995\]](#) and [Amit and Geman \[1997\]](#), [Breiman \[2001a\]](#) proposed an improvement of [RF](#) by introducing the [Bootstrap Aggregation](#) concept of [Breiman \[1996\]](#) which aimed to reduce [variance](#) by making a random selection of samples from the training set to grow each tree in the forest. [Breiman \[2001a\]](#) demonstrated that the generalization error of a forest of tree classifiers depends on the strength of single trees and their correlation [[Koprinska et al., 2006](#)]; the randomness minimizes the correlation to improve accuracy.

Although different extensions and complementary approaches of the original [RF](#) algorithm have been proposed after [Breiman \[2001a\]](#), [Breiman's](#) findings continue being the foundation for most of those works. According to [Pretorius et al. \[2016\]](#), there are two main distinctions between the existing approaches; first, how the independent identically distributed random vectors are obtained [[Tsymbal et al., 2006](#); [Bostrom, 2007](#); [Deng and Runger, 2013](#)]; second, the variety of [ensemble](#) combination strategies and voting schemes [[Robnik-Šikonja, 2004](#); [Tripoliti et al., 2013](#); [Seyedhosseini and Tasdizen, 2015](#)].

2.3.2 Definition

[RF](#) is an [ensemble](#) algorithm that combines multiple [ML](#) algorithms to get a better predictive performance. It works by building multiple decision trees during the training phase and aggregating them to get a unique prediction.

[RF](#) can be used for classification and regression, whose main difference is the type of predicted variable. While in classification, the output variable is categorical as

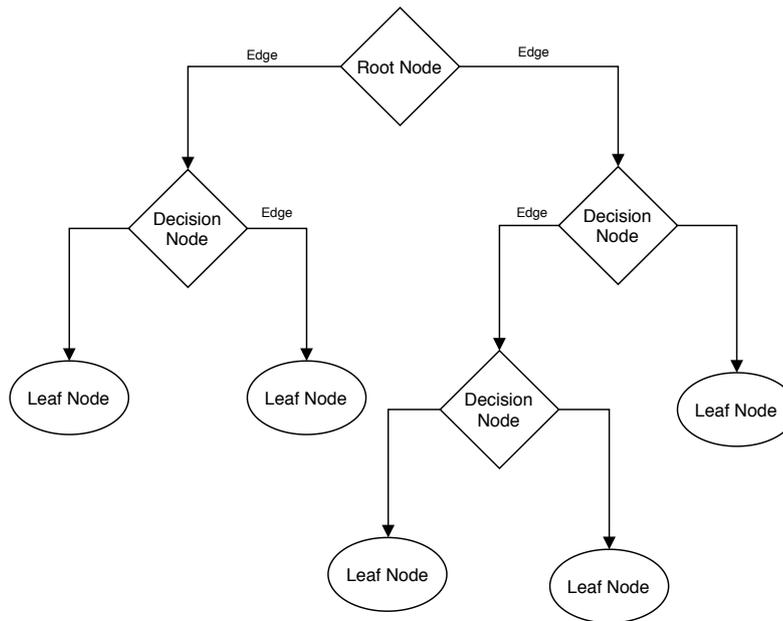


Figure 2.3: Single Decision Tree.

urban/not urban or water/not water, in regression, the variable is a numerical or continuous value as subsidence rate or wind velocity.

To determine the class for unlabeled instances in classification, a simple and effective voting scheme like the majority voting proposed by Lam and Suen [1994] is commonly used. Breiman [2001a] emphasizes that significant improvements in classification accuracy resulted from growing an ensemble of trees and letting them vote for the most popular class.

2.3.3 Single Decision Trees in Random Forest

As previously explained, RF is an ensemble of many decisions trees. A decision tree is a classical supervised learning algorithm that starts at the root node, evaluates the condition, and takes the appropriate edge to the outcome. The process continues until a leaf node represented by class names is found; by then, the object is said to belong to the class named by the leaf.

A decision tree is represented by a tree-graph that expresses a classification rule. The graph is composed of nodes, which are decisions, and edges which are the outcomes or possible paths to the next node (Figure 2.3). There are three types of nodes in the graph:

- The Root Node. It is the top most decision node that performs the first split;
- The Decision Nodes. Those nodes that represent the split in two or more new edges based on a condition and the value of the evaluated feature;
- The Leaf Nodes. They are the terminal nodes that carry the classification.

2.3.4 Main properties of Random Forest

- Integrating multiple models (ensemble method) allows achieving better predictive performance than could be obtained from any of the component models [Fawagreh et al., 2014; Rokach, 2010];
- Using multiple decorrelated trees and random selection of m predictors from a full set of p predictors, reduce variance, and help to overcome the problem

of overfitting. An **overfitted model** does not generalize well to samples that were not encountered during the training;

- By using Mean Decrease Impurity (**MDI**) or Mean Decrease Accuracy (**MDA**) ([Section 3.3.2](#)), **RF** is able to identify which features are the most important to make the predictions;
- **RF** can preserve performance when a large proportion of data is missing and in the presence of outliers. Additionally, it can also handle dataset with high dimensionality;
- **RF** is easy to implement, easily interpretable and it has few parameters to tune;
- **RF** could handle imbalanced datasets through two different approaches: Balanced Random Forest and Weighted Random Forest ([Section 2.5.2](#));

2.4 SUPPORT VECTOR MACHINES

Support Vector Machines (**SVM**) are a family of **ML** algorithms that make a distinction between two classes using a hyperplane or decision surface. The idea is to separate the classes using a decision function that uses support vectors (a selected subset of data points) and optimizes the separation by maximizing the margin around the hyperplane. In a two feature space problem, the decision surface could be seen as a line; in case that the data is not linearly separable, it is possible to transform the problem into a higher spatial dimension.

For a linearly separable problem, it finds the optimal separating hyperplane by maximizing the margin, which is the perpendicular distance across the hyperplane to the closest support vectors on both sides of it ([Figure 2.4](#)). When the data is not separable, two different approaches can be followed: to settle a hyperplane that incurs the least error (soft margin method); or to map the data by conducting a nonlinear transformation using properly chosen basis functions into a higher-dimensional space (kernel trick), where the problem may become linear. [[Dougherty, 2012](#)].

The kernel trick transforms the input to a high dimensional space using a function (kernel function) that allows to express them as inner products in another space. Some well-studied Kernel functions include the Gaussian Kernel, Linear, Polynomial, Multiquadric, Exponential, Laplacian, Bessel, Circular, Hyperbolic Tangent (Sigmoid) Kernel, Inverse Multiquadric, Spherical, Rational Quadratic, Wave, Power, Spline Kernel, B-Spline (Radial Basis Function), Cauchy, and Chi-Square Kernel.

The most used kernels are the polynomial, Gaussian, and Sigmoid Kernels. A Polynomial Kernel has adjustable parameters; slope α , the constant term C , and the polynomial degree d . Gaussian Kernel requires to adjust manually a parameter σ which plays a role in how sensitive it is to noise. For the case of a Sigmoid Kernel, slope α and intercept constant C needs to be manually adjusted.

2.4.1 Main properties of Support Vector Machines

- Solid theoretical and mathematical foundation;
- As the hyperplane fitted is based on an optimization function, **SVM** gives a large space to classify new data offering an accurate generalization;
- **SVM** works well with even unstructured and semi-structured data like text, images, and trees;
- The kernel trick is the real strength of **SVM**; with an appropriate kernel function, it is possible to solve complex problems;

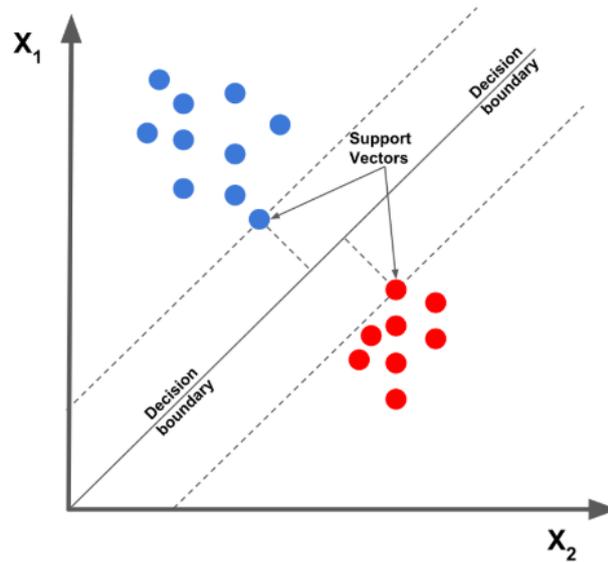


Figure 2.4: Support vectors and classification margin [OpenCv, 2019].

- Complexity on the selection of kernel function parameters. It can be computationally expensive with large datasets.

2.5 IMBALANCED LEARNING

Many real-world applications in ML involve the generation of predictive models using data sets with strongly imbalanced distributions of the target variables [Branco et al., 2016]. Natural hazards detection, diagnosis of rare medical diseases, and fraud detection are among ones of the most critical applications.

For a binary classification problem, a dataset is said to be imbalanced when one of the classes constitutes the majority objects, or it considerably outnumbers the other class. In such cases, the negative class represents the majority, while the positive is the minority. Although any dataset that shows an uneven distribution among classes can be considered imbalanced, in the field of ML, this concept refers explicitly to significant or extreme imbalances.

The imbalanced distribution of datasets is a big challenge for supervised learning. As most of the ML algorithms aim to minimize the overall error rate rather than focus on the positive or target class [Chen et al., 2004], the prediction will be biased towards the majority class. Therefore, classifiers tend to give an extremely imbalanced degree of accuracy, having a majority class with accuracies close to 100% and a minority class with accuracies around 0-10% [He and Garcia, 2008].

2.5.1 Handling imbalanced datasets

Handling imbalanced datasets is influenced by different factors, including the particular case study, the chosen supervised algorithm, and the ratio of imbalance. Japkowicz and Stephen [2002] conducted a research to evaluate the nature of the class imbalance problem and found that the smaller the overall size of the training set is, the greater the effect of class imbalance will be.

Several researchers have discussed different strategies to address the problem of imbalanced datasets [Chen et al., 2004; Japkowicz and Stephen, 2002; He and Garcia, 2008; Branco et al., 2016]. Two of the most commonly used approaches are:

1. Sampling Methods. They consist in the modification of the original dataset using some techniques that allow either down-sample the majority class, over-sample the minority class, or both (hybrid methods) [Chen et al., 2004; He and Garcia, 2008]. Some of the most well-known strategies include random under/over-sampling, data cleaning approaches, distance techniques, clustering algorithms, and synthesizing new data from existing observations [Branco et al., 2016].
2. Cost-Sensitive Methods. They consist in modifying the relative cost linked with misclassification of the classes. For instance, it will assign a high cost to misclassification of the minority class [Chen et al., 2004] to compensate for the imbalance ratio between the classes; thus, with a ratio of imbalance of 1:50, the cost of misclassifying an observation of the minority class is set to 50 times that of misclassifying an observation of the majority class [Japkowicz and Stephen, 2002].

2.5.2 Random Forest for imbalanced datasets

Similar to most of the classification algorithms, RF is biased toward the majority class as it tries to minimize the overall error rate. Chen et al. [2004] proposed two different approaches to address the challenge of imbalanced classification in RF:

1. Weighted Random Forest (WRF). This approach is based on the cost-sensitive methods (Section 2.5.1). The algorithm adds more weights to the minority class, thus assigning a higher penalty when misclassifying the minority. The class prediction of each terminal node is defined by a weighted majority vote, and a final prediction is determined by aggregating all votes from each tree [Chen et al., 2004].
2. Balanced Random Forest (BRF). It is a variant of RF that combines the down-sampling majority class method and ensemble learning by artificially altering the class distribution to have an equal number of the classes at each tree [Chen et al., 2004].

2.5.3 Metrics for model performance

Assessing the model performance is a crucial step when applying any supervised classification method. A well-known representation of the model results is the confusion or error matrix where actual values are compared with the predicted values. In this representation, the total number of rows (actual class) equals the number of columns (predictive class) so that for a given number of k classes, the matrix dimension will be $k \times k$ [Lemmens, 2011, 1994].

For a binary classification problem with a positive and negative class, the confusion matrix is given by a 2×2 table (Table 2.2).

	Predictive Class	
Actual Class	Negative	Positive
Negative	Tn	Fp
Positive	Fn	Tp

Table 2.2: Confusion matrix for binary classification. Tp are the true positives predictions, Tn are the true negatives, Fp are the false positives, and Fn are the false negatives

One of the most commonly used metrics derived from the confusion matrix is the overall accuracy. It is defined as the ratio between the number of correct predictions and the total number of observations (Equation 2.1):

$$Accuracy = \frac{\mathcal{T}_p + T_n}{\mathcal{T}_p + T_n + F_p + F_n} \quad (2.1)$$

Where \mathcal{T}_p are the true positives or correctly classified true values, T_n are the true negative or correctly classified false values, F_p are the false positives or incorrectly classified true values, and F_n are the false negative or incorrectly classified false values.

Despite the popularity and simplicity of the overall accuracy as a metric for estimating a classifier performance, for imbalanced datasets where often the minority is the positive class and the majority the negative, it is not an adequate measure as the ML algorithm will tend to favor the majority class [Chen et al., 2004]. Since the overall accuracy is very sensitive to the data distribution [He and Garcia, 2008], a large number of true negative predicted observations will considerably increase the classification accuracy results; therefore, the impact of the minority class (target class) is reduced when compared to that of the majority class [Branco et al., 2016].

Alternative model evaluation metrics for imbalanced classification have been described by previous works [Chen et al., 2004; He and Garcia, 2008; Branco et al., 2016]. Similar to the overall accuracy, those metrics are also functions of the confusion matrix. Some of the most commonly used are:

1. Precision. It measures how many correct predictions are made from a total number of positively predicted observations (Equation 2.2). It is a useful metric when the cost of false positives is high; thus, the higher the precision, the less the number of false positives predicted.

$$Precision = \frac{\mathcal{T}_p}{\mathcal{T}_p + F_p} \quad (2.2)$$

2. Recall. It measures how many correct predictions are made from the total amount of observations (Equation 2.3). In contrast to precision, its applicability has more significance when the cost of false negative is high. Recall is also known as true positive rate (TP_{rate}), sensitivity or probability of detection.

$$Recall = \frac{\mathcal{T}_p}{\mathcal{T}_p + F_n} \quad (2.3)$$

3. F-Measure. It combines precision and recall to assess the performance of the classification in terms of a ratio of the weighted importance on either recall or precision [He and Garcia, 2008]. It can be calculated using Equation 2.4:

$$\mathcal{F}_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (2.4)$$

Where β is a coefficient to adjust the relative importance of recall with respect to precision. For a value of $\beta = 1$, precision and recall have the same weight.

In contrast to using precision and recall separately, \mathcal{F}_β gives more insight about the effectiveness of a classifier on correctly making predictions [He and Garcia, 2008; Branco et al., 2016]. According to Equation 2.4, \mathcal{F}_β will be high when both precision and recall are high.

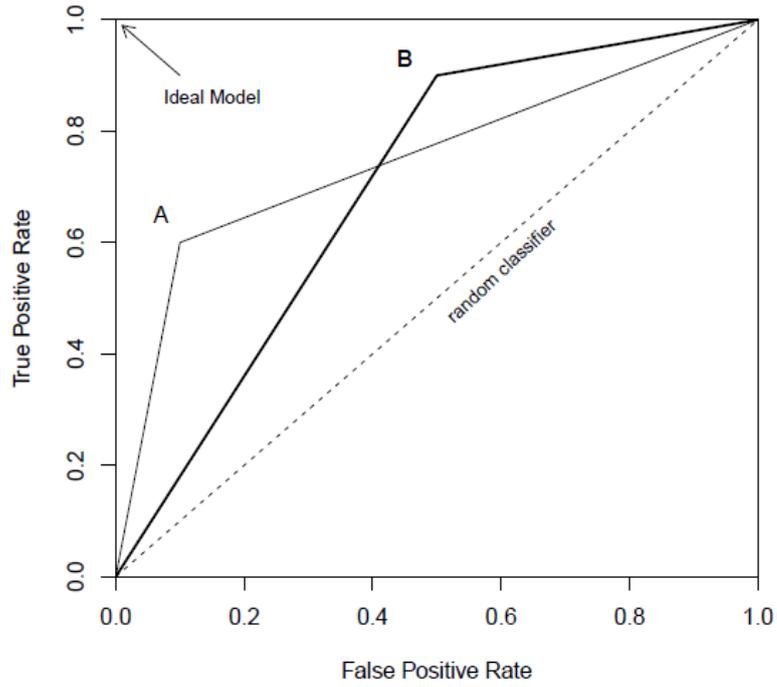


Figure 2.5: ROC curve of three classifiers: A, B and a random classifier (Branco et al. [2016]).

- Receiver Operating Characteristic (ROC) curve. It consists in a graphical representation of the relative trade-off between the false positive rate (FP_{rate}) (Equation 2.5), and the TP_{rate} (Equation 2.3) or sensitivity at different thresholds settings:

$$\mathcal{F}P_{rate} = \frac{F_p}{F_p + T_n} \quad (2.5)$$

The ROC curve is generated by plotting TP_{rate} on the y-axis against the FP_{rate} on the x-axis. Every point on the curve corresponds to a different value of a decision/threshold parameter used to classify an example as belonging to the positive class [Branco et al., 2016]. In general terms, it gives insight about the degree of a classifier to distinguish between the two classes.

- Area Under the Curve (AUC). It can be defined as the area under the ROC curve and evaluates how much a model can differentiate between different classes. The AUC can be computed using Equation 2.6:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (2.6)$$

3

RESEARCH METHODOLOGY

This chapter describes the methodology proposed to address the main research questions of this thesis (Section 1.2). Concepts, theories, algorithms, and mathematical foundations linked to the stages illustrated in Figure 1.1 are discussed in this chapter.

The overview of the methodology workflow is introduced in Section 1.5. Three main stages are defined: (a) pre-processing (b) image segmentation, and (c) image classification (Figure 1.1).

3.1 IMAGE PRE-PROCESSING

3.1.1 Samples set preparation

The sample preparation involves the review of previous landslides inventories and the manual identification/digitization of landslides events or potential training/testing samples. As a complete and accurate landslide inventory for a specific study area is not available on open databases, we decided to generate worldwide ground truth samples through heads-up digitization of landslide events using Sentinel-2 times series and DEM. The following steps are considered to build the sample datasets:

- Filter large-catastrophic scale landslides out from previous landslide inventories (National Norwegian landslide inventory ¹, NASA Catalog ², the New Zealand's National Landslide Database ³). Based on the GSD and temporal coverage of the input datasets, additional filters are included: high location accuracy (< 1,000m) and event date from August 2015 and onwards;
- Apply previous knowledge in geomorphology and photogeology to interpret the geomorphic expression of landslides using Sentinel-2 time series;
- Use hill-shading images derived from DEM to visualize slopes and support the identification of the potential landslides areas;
- Use visual change on vegetation to identify fresh landslide scarps;
- Use social media and online resources to support the identification of landslides occurrences not included in previous databases.

3.1.2 Image set preparation

The image set preparation consists in generating cloud-free images collections of pre- and post-landslides events.

The main limitation when handling satellite imagery from different scenes is the scarce availability of pixels with non-null spectral information in the presence of seasonal changes in vegetation, clouds, and shadows. The challenge is to apply the most convenient method to remove pixel signatures that are not related to the actual

¹ <https://www.skredregistrering.no/Forsiden>

² <https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog/h9d8-neg4>

³ <https://data.gns.cri.nz/landslides/>

reflectance of the target objects and generate cloud-free images without altering their original properties.

Three main strategies are considered to select the most appropriate images:

1. Cleanest image within a period. It consists in finding the least clouded pre- and post-landslide event images within a pre-defined period. The initial search is conducted within four months (before and after the event occurrence) in order to avoid seasonal changes; however, depending on the case study, the number of months could be modified.
2. First cloud-free image. It consists in getting the first pre- and post-event image with a cloud percentage less than a pre-defined threshold (approximately 30%).
3. Image composition. It is based on optimizing pixels spectral signature or identifying the best-suited observation within a pre-defined temporal window. The appropriate composition strategy depends on the cloudiness of the region and the sensor revisit time. As clouds often have a high spectral reflectance, the composite is created using bands median values (per pixels) from all images under consideration. A disadvantage of this strategy is that it may be affected by land cover changes or new event occurrences over the selected period. While short periods may result in occurrences of data gaps, long periods will increase the amount of valid data but may result in visible changes of vegetation or land cover.

As shown in [Figure 3.1](#), different pre-processing steps can be applied to handle satellite images distributed on a global scale. A pre-processing algorithm ([Algorithm 3.1](#)) to retrieve cloud-free images pre- and post landslide occurrence is developed based on the three strategies described above. To find the most appropriate image, the algorithm analyzes large time-series of Sentinel-2 imagery using as input the geographical coordinates and the landslide date of occurrence. All the operations related to the computation of [landslides diagnostic features](#) at pixel level are conducted within the same algorithm, including the band ratioing, image normalization, and image difference ([Section 3.1.3](#)). After retrieving the cloud-free images and calculating the features, the images are downloaded using a fix bounding box of 5km x 5km ([Figure 3.2](#)).

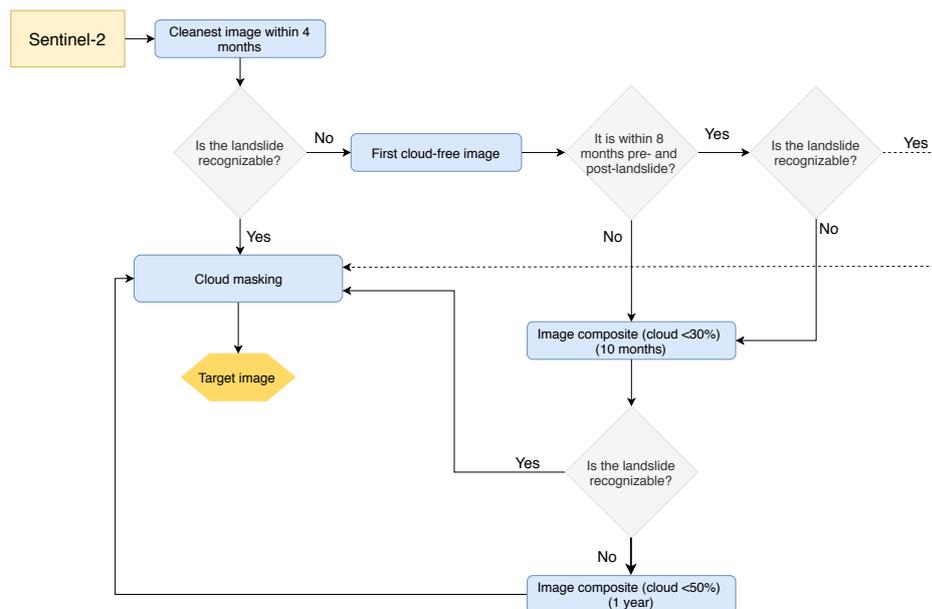


Figure 3.1: Cloud-free images pre-processing workflow.

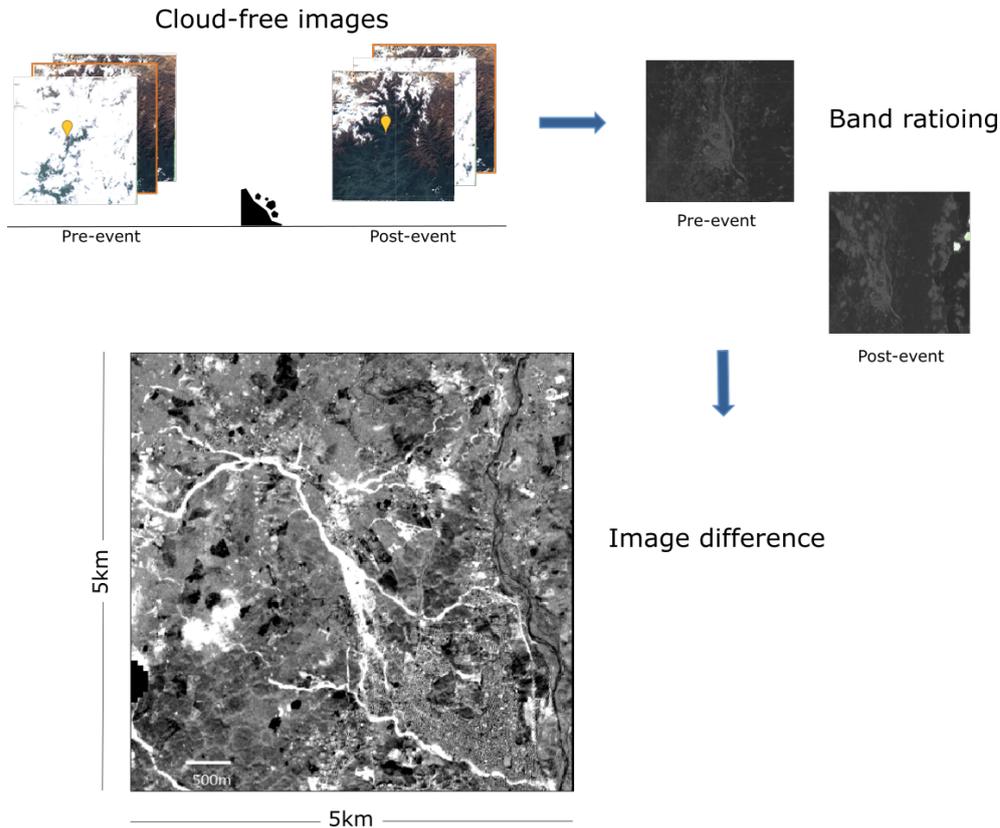


Figure 3.2: Image set preparation.

Algorithm 3.1: Landslide Image Pre-processing (GEE)

Input: satellite, landslides

- 1 Initialize image collector
- 2 **for** *landslides* **do**
- 3 Add landslide to new before Processor
- 4 Add landslide to after Processor
- 5 **Process Before (landslide):**
- 6 calculate event time range before event
- 7 check for cloud free images in time range
- 8 apply cloud masking
- 9 add to image collection
- 10 **Process After (landslide):**
- 11 calculate event time range
- 12 check for cloud-free images in time range after event
- 13 apply cloud masking
- 14 add to image collection
- 15 **Process Collector:**
- 16 **if** *image collection is ready* **then**
- 17 calculate band rationing
- 18 normalize the image bands Red,Green,Blue
- 19 calculate NDVI
- 20 calculate Brightness
- 21 calculate image difference for NDVI , Brightness, band rationing
- 22 download image difference as .tiff
- 23 **end**

Cloud masking

As seen in [Figure 3.1](#), a cloud masking algorithm is applied to remove clouds that remain after an appropriate cloud-free image is selected. The applied algorithm first selects the Quality Assessment (QA60) band and uses bit 10 and bit 11 to check whether or not dense or cirrus clouds are present; if they exist, the entire pixel values are set to "0" using a masking operation. Sentinel-2 level-1C includes the QA60 band with one bitmask band containing cloud mask information for both, dense (bit 10) and cirrus clouds (bit 11). For more details about Sentinel-2 program and the QA60 bands, the reader is referred to [Section 4.1](#).

3.1.3 Features computation at pixel level

Spectral and spatial information are extracted from Sentinel-2 imagery and ALOS World 3D-30m DEM (AW3D30) (see [Figure 1.1 a](#)). From the extracted pixel signatures, potential [landslides diagnostic features](#) are calculated using band ratioing and image difference techniques ([Table 3.1](#)).

The initial selection of features is based on previous works on landslide detection ([Section 2.2](#)) and [knowledge-based features](#) commonly used by experts to identify landslides during visual image interpretation. To provide the basis to develop object-based landslide mapping routines, [Martha et al. \[2010\]](#) updated and synthesized the diagnostic features for semi-automatic detection (recognition and classification) of landslides. In addition to [spectral features](#) (e.g. NDVI, Brightness), [landslides diagnostic features](#) can include [topographic features](#) from DEM. Previous researchers have attempted to quantify some of those features. [Pike \[1988\]](#) calculated the geometric signature derived from DEM for a set of topographic variables that separates a landslide from its surroundings. [Iwahashi and Pike \[2007\]](#) used slope gradient and terrain texture for automatic classification of topography. [McDermid and Franklin \[1994\]](#) demonstrated that the integration of spectral and topographic features produces better results than working with only spectral information.

Feature nature	Feature	Target Image	Source
Spectral	NDVI	post-event	Sentinel-2
Spectral	GNDVI	post-event	Sentinel-2
Spectral	Brightness	post-event	Sentinel-2
Spectral	RGD	pre-/post-event	Sentinel-2
Spectral	VID	pre-/post-event	Sentinel-2
Spectral	BrightnessD	pre-/post-event	Sentinel-2
Textural	NDVI _{texture}	Post-event	Sentinel-2
Spatial	Slope	pre-event	ALOS (2011)

Table 3.1: Initial landslide diagnostic features.

Band ratioing technique

Band ratioing is a powerful remote sensing technique whose primary purpose is to highlight the anomaly of the target object while suppressing irrelevant features. Every object has a spectral reflectance pattern in different wavelength portion [[San et al., 2004](#)]. Depending on the analyzed object, it will tend to absorb in one band and reflect in another band; those key spectral signatures are used in remote sensing for different geosciences and environmental applications, such as minerals exploration, vegetation monitoring, and shadows detection.

RED/GREEN RATIO. The Red/Green ratio (**R/G**) helps to minimize the effect of brightness variation when applying image difference techniques. Identical surface materials can exhibit different brightness values caused by topographic slope and aspect, shadows, seasonal changes, and differences in sunlight illumination angle and intensity [Haque, 2017].

The method is applied to single multi-spectral Sentinel-2 images (pre- and post-event) using the Red and the Green bands. For each pixel, the reflectance value of the Red band is divided by the value of Green band (Equation 3.1):

$$rg(i, j) = \frac{gk(i, j)}{gl(i, j)} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.1)$$

Where $rg(i, j)$ is the output **R/G** value for pixel x at row i and column j , $gk(i, j)$ is the reflectance value for the Red band at row i and column j , and $gl(i, j)$ is the reflectance value for the Green band at row i and column j .

NORMALIZED DIFFERENCE VEGETATION INDEX. The **NDVI** is often used in remote sensing for monitoring vegetation. According to Bannari et al. [1995], it exhibits better sensitivity than individual spectral bands for green vegetation identification. Detecting sudden changes in **NDVI** is a common practice to analyze disturbances on vegetation as well as for rapid landslide identification [Yang et al., 2012].

NDVI is built on the fact that chlorophyll largely reflects the Near-Infrared (**NIR**) band and absorbs the Red band [Lemmens, 2011]. According to Bannari et al. [1995], **NDVI** values approximate to 1 in healthy vegetation, to 0 in bare ground, and to -1 in water bodies. The mathematical expression used to calculate the **NDVI** is shown in Equation 3.2:

$$ndvi(i, j) = \frac{gq(i, j) - gk(i, j)}{gq(i, j) + gk(i, j)} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.2)$$

Where $ndvi(i, j)$ is the output **NDVI** value for pixel x at row i and column j , $gq(i, j)$ is the reflectance value for the **NIR** band at row i and column j , and $gk(i, j)$ is the reflectance value for the Red band at row i and column j .

From the **NDVI** values, a textural feature ($NDVI_{texture}$) that uses the **NDVI** standard deviation within a neighbor is computed using a sliding window (kernel=8 neighbors) over the input image. $NDVI_{texture}$ aims to highlight differences in texture induced by spatial changes in vegetation.

GREEN NORMALIZED DIFFERENCE VEGETATION INDEX. The Green Normalized Difference Vegetation Index (**GNDVI**) is a vegetation index widely used in many phenological applications for being more sensitive than **NDVI** to chlorophyll concentration changes. In the context of landslide detection, previous researchers [Hölbling et al., 2015; Veena et al., 2016] have used **GNDVI** to detect small-scales vegetation changes not captured by **NDVI** and to eliminate false positive as barren lands.

GNDVI is computed similarly than **NDVI** (Equation 3.2), but the Red band is replaced by the Green band (Equation 3.3):

$$gndvi(i, j) = \frac{gq(i, j) - gl(i, j)}{gq(i, j) + gl(i, j)} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.3)$$

Where $gndvi(i, j)$ is the output **GNDVI** value for pixel x at row i and column j , $gq(i, j)$ is the reflectance value for the **NIR** band at row i and column j , and $gl(i, j)$ is the reflectance value for the Green band at row i and column j .

Image normalization

Image normalization is a technique applied before calculating the Vegetation Index Difference (VID) and Brightness Difference (BrightnessD). This method is useful to reduce atmospheric and radiometric differences between the pre- and post-event image. Singh [1989] states that the use of such indices can minimize the topographic effects and differences in sun illumination when working with multi-temporal images.

Hölbling et al. [2015] applied a normalization technique to reduce the bias introduced by absolute values of NDVI when comparing pre- and post-event images for detecting landslides. For this study, a similar normalization approach than the one proposed by Hölbling et al. [2015] and Singh and Gupta [2014] is used. It consists in calibrating the spectral values of the pre-event image to the radiometric characteristic of the post-image (Equation 3.4).

$$gx_n(i, j)_{t_1} = gx(i, j)_{t_1} \frac{\mu_x(t_2)}{\mu_x(t_1)} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.4)$$

Where t_1 is the first date (pre-event), t_2 is the second date (post-event), $gx_n(i, j)$ is the normalized value of pixel x at row i and column j in the Red, Green, Blue, or NIR bands, $gx(i, j)$ is the absolute value of pixel x at row i and column j in the Red, Green, Blue, or NIR bands, and μ_x is the mean values (image global mean) of all pixels of the image for Red, Green, Blue, or NIR bands.

Previous works [Collins and Woodcock, 1994; Coppin and Bauer, 1996; Song et al., 2001] have demonstrated that relative radiometric corrections to normalize intensities bands of multitemporal imagery are suitable for different change detection applications; therefore, complex corrective algorithms not necessarily lead to a significant increase in change detection accuracy [Chen et al., 2012]. Hölbling et al. [2015] states that using image objects instead of pixels decreases the effects of radiometrically and atmospherically not perfectly corrected pixel values, thus making the task of atmospheric and radiometric image correction not as critical as it is for pixel-based change detection.

Image difference

The image difference method [Singh, 1989] is applied to detect changes before and after the landslide occurrence. With this technique, images from t_1 (pre-event) and t_2 (post-event) are subtracted, pixel by pixel to produce a further image which represents the change between the two times. The technique detects a change in reflectances or radiances values so that those differences are taken as a measure of change [Mahmoodzadeh, 2007].

According to Ingram et al. [1981], variation in spectral signature caused by land cover changes (or changes by landslides) should be larger than the changes induced by other factors, including differences in illumination, atmospheric conditions, sensor calibration, and differences in soil moisture.

RED/GREEN RATIO DIFFERENCE. The Red/Green Difference (RGD) is computed using the Red/Green ratio (R/G) previously calculated with Equation 3.1 for both pre- and post-event images (Equation 3.5):

$$rgd(i, j) = rg(i, j)_{t_2} - rg(i, j)_{t_1} + c \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.5)$$

Where $rgd(i, j)$ is the output RGD value for pixel x at row i and column j , $rg(i, j)$ is the pixel value for the R/G feature at row i and column j , t_1 is the first date (pre-event), t_2 is the second date (post-event), and c is a constant to produce positive values.

VEGETATION INDEX DIFFERENCE. The **VID** is calculated using **NDVI** values from normalized pre-event image and absolute **NDVI** values from post-event image (Equation 3.6):

$$vid(i, j) = ndvi_n(i, j)_{t_1} - ndvi(i, j)_{t_2} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.6)$$

Where $vid(i, j)$ is the output **VID** value at row i and column j , $ndvi_n(i, j)$ is the normalized value of **NDVI** at row i and column j , $ndvi(i, j)$ is the absolute value of **NDVI** at row i and column j , t_1 is the first date (pre-event), and t_2 is the second date (post-event).

BRIGHTNESS AND BRIGHTNESS DIFFERENCE. The Brightness is calculated for pre- and post event images using Equation 3.7:

$$Br = \frac{1}{gk(i, j) + gl(i, j) + gm(i, j)} \quad (i, j) : i = 1, \dots, n; j = 1, \dots, m \quad (3.7)$$

Where Br is the brightness value for pixel x at row i and column j , $gk(i, j)$ is the pixel value for the Red band at row i and column j , $gl(i, j)$ is the pixel value for the Green band at row i and column j , and $gm(i, j)$ is the pixel value for the Blue band at row i and column j .

Areas affected by landslides occurrences are commonly related to high Brightness values due to loss of vegetation and exposure of fresh rock and soil [Martha et al., 2010]. Very high and low values of Brightness relative to the landslides signature are used in this study to detect and remove clouds and shadows respectively.

The **BrightnessD** feature or difference in Brightness is computed using similar equation than Equation 3.6, but replacing **NDVI** by Brightness values.

Topographic feature

The slope or steepness of the ground surface is extracted from **AW3D30 DEM**. It is calculated within Google Earth Engine (**GEE**) API using a moving window of 3x3 kernels and the 4-connected neighbors (horizontal and vertical neighbors of the central pixel) of each pixel.

3.2 IMAGE SEGMENTATION

As explained in Section 1.5, a two-step segmentation approach is devised to address the challenge of multi-scale objects (landslides of different sizes) using exclusively open and free-of-charge resources. This strategy consists in an image segmentation stage that segments the image using a k-means implementation, and a merging algorithm using region growing to balance the dataset (see Figure 1.1 b).

3.2.1 Initial Segmentation

The initial segmentation aims at creating primitive homogeneous segments using the **RGD** computed with Equation 3.5. **RGD** is selected as the input feature for segmentation as it is the homogeneity criterion that demonstrated better performance to accurately segment landslides of different sizes and shapes, and with a wide range of **spectral features** values.

The image segmentation is conducted using the operational large-scale segmentation algorithm proposed by Shepherd et al. [2019] for exclusively clustering remote

sensing data. The algorithm uses a k-means implementation for seeding and provides support for a minimum mapping unit through an innovative iterative elimination process. Shepherd et al. [2019] demonstrated that their algorithm has comparable performance to the multi-resolution segmentation approach [Baatz, 2000] implemented within eCognition software. As the initial image segmentation relies on detailed information of pixels (e.g. RGD, NDVI), the Shepherd et al. [2019] algorithm is a suitable approach due to its effectiveness and simplicity in parameter tuning across a large variety of scenes, and its low-computational cost for large geographical areas. Additionally, the algorithm does not require any *a priori* information about the study area and allows to work with landslides of different sizes, including small-scale events occurring across different land covers, landscapes, and geomorphological settings.

The k-means implementation used within the Shepherd et al. [2019] algorithm is a simple and widely used clustering algorithm that iteratively partitions a given dataset (e.g. remotely sensed image) into a user-specified number of clusters k [Wu et al., 2008; Dougherty, 2012; Khorram et al., 2016]. The algorithm aims at minimizing the sum of distances from each pixel to its k centroid over all clusters in the image [Dhanachandra et al., 2015].

The following workflow summarizes the steps for segmenting an image using k-means algorithm [Dhanachandra et al., 2015], and includes the additional steps 6-8 of Shepherd et al. [2019] algorithm:

1. Initialize k centroids at random rgd values in a one-dimensional space.
2. Calculate the Euclidean distance d between the centroids and each pixel on the image using Equation 3.8:

$$d = | p_{rgd} - c_k | \quad k = 1, \dots, k_{max} \quad (3.8)$$

Where d is the distance between the pixel and the centroid, p_{rgd} is RGD value for pixel p , and c_k is the centroid of the cluster.

3. Assign all the pixels to the nearest centroid based on distance d .
4. After assigning all pixels, recalculate the new position of the centroid using Equation 3.9:

$$c_k = \frac{1}{n_k} \sum_{p \in k} p_{rgd} \quad k = 1, \dots, k_{max} \quad (3.9)$$

Where c_k is the centroid of cluster k , n_k is the total number of pixels p in cluster k , and p_{rgd} is the RGD value for pixel p .

5. Repeat the process until it satisfies a certain tolerance and no cluster assignment changes.
6. Convert the cluster pixels into an image. The k-means classified pixels are clumped to define a set of geographically uniquely-labeled regions.
7. Iteratively eliminate segments if they are below the minimum mapping unit threshold to the neighboring segment that is spectrally closest (regions are eliminated based on size).
8. Relabel segments to ensure they are sequentially numbered.

The two key input parameters required to apply the Shepherd et al. [2019] are the initial number of seeds k (number of clusters in the feature space) and the minimum segment size (measured as the number of pixels) for elimination. k is

estimated using the Elbow method ⁴ and through experimentation, while the minimum segments size (measured number of pixels) is controlled by the minimum size of the target landslides (100px) (see Section 1.4).

Elbow method

Despite the advantages of k-means in terms of its simplicity and scalability, the quality of the final clustering depends on the initial number of k , which can not be directly estimated. The Elbow method is one of the main techniques used to estimate this optimal number of k . The basic idea of this method is to fit the model with a range of values of k and minimize the total sum of squared errors over all the clusters.

The main steps to estimate the optimal number of k using the Elbow method are:

1. Compute k-means for different values of k .
2. Calculate the sum of squared errors WSS using Equation 3.10:

$$WSS = \sum_{p \in k} (p_{rgd} - c_k)^2 \quad (3.10)$$

Where WSS is the sum of square errors between each pixel value and its group mean, p_{rgd} is the RGD value for pixel p , and c_k is the centroid of cluster k .

3. Plot the WSS against the number of clusters k .

The parameter to be minimized corresponds to total WSS over all clusters. Generally, the first clusters add much information, and at some value of k , the gain drops dramatically and gives a point of inflection on the plot (the elbow) that is considered as a good indicator of the optimal number of clusters [Bholowalia and Kumar, 2014].

As the input dataset of this study consists of several satellite images, the average of all k values is considered as an initial estimation of k . This value is adjusted to a higher value through experimentation in order to enable the proper segmentation of the smallest target objects (landslide $\approx 10,000m^2$) (Section 5.2).

For this study, the oversegmentation approach is chosen rather than undersegmentation. While a low value of k merges small-scale landslides with larger segments, a higher value of k leads to oversegmentation but preserved landslides segments of different sizes. According to Carleer et al. [2005], all target objects in an image can not be extracted with only one segmentation step. A certain degree of oversegmentation is generally accepted rather than undersegmentation, as merging strategies can be later applied to objects of the same class [Shepherd et al., 2019]. In contrast, undersegmentation can not be recovered during classification as target objects have been already misidentified or merged with larger segments [Debeir, 2001; Carleer et al., 2005] and consequently, spatial and contextual information is lost.

3.2.2 Merging algorithm

Oversegmentation leads to a high-imbalanced dataset with the *landslide* or target class being the minority; thus, resulting in poor model performance. To balance the dataset, *non-landslide* segments are merged (down-sampling, Section 2.5) using an implementation of the region growing algorithm at segment level. This strategy generates larger and fewer *non-landslide* segments that preserve a degree of homogeneity in terms of their *NDVI* values. The *NDVI* is selected as the first criterion of homogeneity as it has been successfully used in previous works to discriminate

⁴ [https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))

landslides (low NDVI) from vegetated areas (Equation 3.2). Equation 3.11 is used to calculate the measure of similarity:

$$Sm = | \overline{ndvi}_{ws} - ndvi_{nb} | \quad (3.11)$$

Where \overline{ndvi}_{ws} is the weighted mean of the seed region (Equation 3.12), and $ndvi_{nb}$ is the NDVI of the neighbor. A neighbor segment is merged if $S_m < t$, with t being a set threshold.

The main challenge of the homogeneity criterion is to define a suitable threshold that could be applied to different NDVI distributions. For this, a decision algorithm that uses a threshold adapted to the statistical distribution of spectral features is devised (Algorithm 3.2). It considers the non-local or global NDVI, VID, and RGD means to merge similar segments with low probability of being landslides. *landslide* segments tend to be located on the left side of the NDVI distributions; therefore, using the non-local NDVI mean, it is possible to merge all segments located on the right side of the distribution without the risk of including the *landslide* segments. With this approach applied to Gaussian distributions (Figure 3.3), it is possible to safely merge 50% of the segments without the risk of class mixture.

To merge the remaining *non-landslide* segments, an additional criterion based on VID and RGD is applied. As *landslide* are directly linked to changes before and after their occurrences, they tend to have higher values of VID and RGD compared to the image mean; therefore, segments located to the left side of the distribution can be safely merged. Finally, the remaining segments are merged using a fine-grained or non-flexible homogeneity threshold. With the NDVI, VID, and RGD criteria and the similarity measure (S_m), it is possible to merge a large number of *non-landslides* avoiding the class mixture and reducing the probability to merge landslides.

Algorithm 3.2: Decision based on homogeneity criterion.

Input: region, neighbor, NDVI non-local mean, NDVI non-local std, VID non-local, VID non-local std, RGD non-local

Output: True/False

```

1 set coarse-grained threshold
2 set fine-grained threshold
3 calculate similarity criteria ( $S_m$ ) using NDVI
4 right VID boundary = vid mean + 2 std
5 right RGD boundary = rgd mean + 2 std
6 if NDVI neighbor > NDVI non-local mean AND  $S_m <$  coarse-grained threshold
   then
7   return True
8 if VID neighbor < right boundary VID AND  $S_m <$  coarse-grained threshold
   then
9   return True
10 if RGD neighbour < right RGD boundary AND  $S_m <$  coarse-grained threshold
    then
11   return True
12 if  $S_m <$  fine-grained threshold then
13   return True
14 return False
```

The implemented region growing algorithm starts by iteratively selecting seed segments (segments from which the region starts to grow) and finding their nearest neighbors using the K-Nearest Neighbor (KNN) algorithm (Figure 3.4). The chosen seeds are those segments with the lowest ratio of change pre- and post-landslides

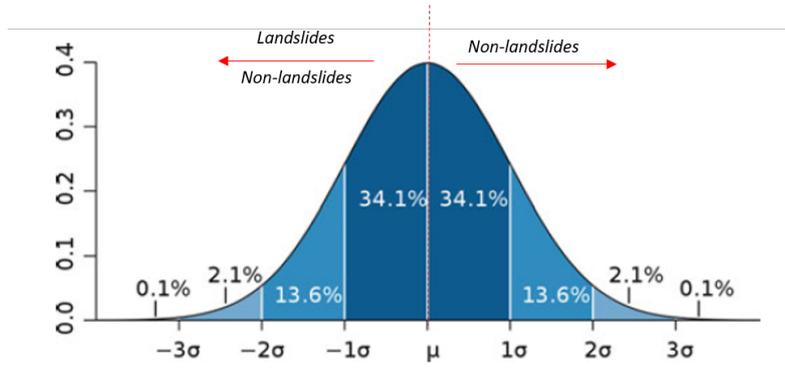


Figure 3.3: Gaussian Distribution function (modified from Dougherty [2012]).

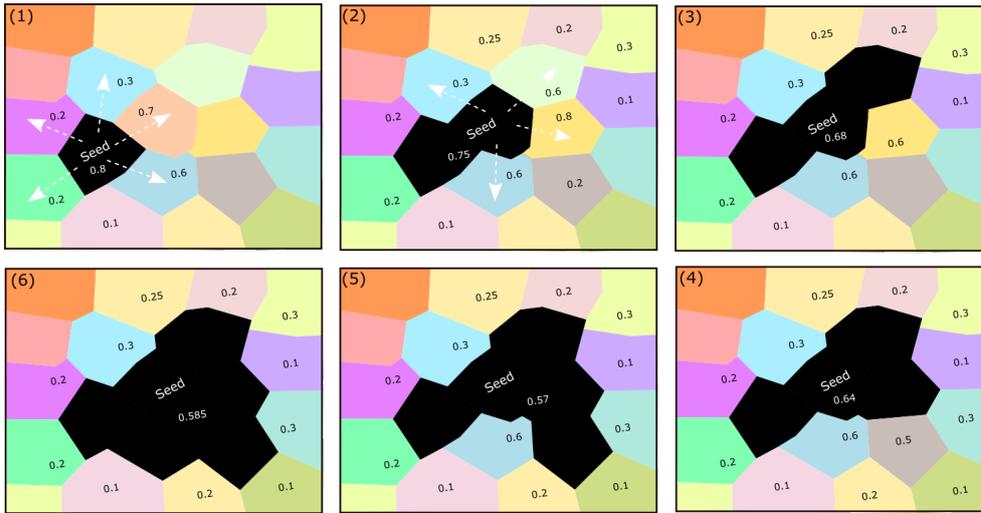


Figure 3.4: Region growing representation at segment level. The numerical values are hypothetical NDVI values. (1) The algorithm finds the five nearest adjacent neighbors to the seed. (2) The region starts to grow from the seed; the first adjacent segment to be compared is the closer in similarity (closest in NDVI); if the segments satisfy the homogeneity criterion, then the adjacent segment is added to the seed and the NDVI is recalculated using the weighted mean. (3) The algorithm finds the adjacent neighbors of the added segment. (4) Next segment closer in similarity is added to the region, and the NDVI is recalculated. (5), (6), the process is repeated until a stop is reached and no new segment can be added.

occurrences (minimum *RGD*). With this criterion, there is a high probability that the *non-landslide* segments are merged first.

The first adjacent neighbor to be compared is the closest in similarity to the seed region. If the segments are similar enough to satisfy the merging condition, then the segment is added to the region and the *NDVI* of the new seed region is recalculated using Equation 3.12:

$$\overline{ndvi}_{ws} = \frac{\sum_{i=1}^n w_i ndvi_i}{\sum_{i=1}^n w_i} \quad (3.12)$$

Where \overline{ndvi}_{ws} is the weighted mean of the seed region, w_i is the segment weight (segments area), and $ndvi_i$ is the *NDVI* value of segment i . The remaining *landslides diagnostic features* are recalculated using a similar expression than Equation 3.12, but replacing the *NDVI* value by the corresponding feature value.

The growing continues by examining the adjacent neighbors of the added segment. If one of the adjacent neighbors of the newly merged segment is also adjacent

to the initial seed region, the segment is ignored. The process is then repeated until it reaches a stop, meaning that not more joinable segments remain (Figure 3.4, step 6).

Once all neighbors of the first seed are evaluated, the algorithm moves to the following seed and iteratively repeats the whole process until no seeds are available. The pseudo-code for this implementation is presented in Algorithm 3.3.

Algorithm 3.3: Region growing at segment level

Input: list of segments (Kd-tree data structure)

Output: list of regions

```

1 while list of segments is not empty do
2   seed = []
3   region = []
4   find the segment with minimum RGD
5   add the segment to the seeds list
6   add the segment to the regions list
7   remove segment from the segments list
8   while seeds do
9     find the neighbors of every seed segment;
10    for every neighbor do
11      if similarity measure gives true then
12        add current neighbor to the region;
13        remove current point from the segment list;
14        add neighbor to the seeds list
15 return list of regions

```

Algorithm optimization

For optimization purposes, a combined approach of **KNN** and K-Dimensional tree (**KD-tree**) data structure is implemented. **KNN** is a simple and commonly used algorithm for information retrieval. The algorithm works as follows: for every target sample (e.g. points or segments), it calculates the Euclidean distance to every other sample in the dataset. The neighbors are the k samples that are closest to the target.

When using **KNN** it compares all the samples in the dataset with the target. To optimize this searching operation, a **KD-tree** structure is generated to store the segments efficiently. This spatial access method supported the optimal selection of adjacent segments based on their spatial location. **KD-tree** splits the global space into sub-spaces by repeatedly selecting a sample in the dataset. van Oosterom [1999] describes the algorithm as follows: the root of the tree corresponds to the global space of interest. Every internal node of the **KD-tree** is a k -dimensional sample that also corresponds to a sample on the rectangular region (Figure 3.5). The rectangular region is divided into half by the x -coordinate of the stored sample on the odd levels and by the y -coordinate on the even levels in the tree.

3.2.3 Features computation at segment level

In addition to the initial **landslides diagnostic features** depicted in Table 3.1, contextual information is introduced by computing the value of the features relative to the information contained in the image (Equation 3.13):

$$f_{deviation} = f_i - \overline{f_w} \quad (3.13)$$

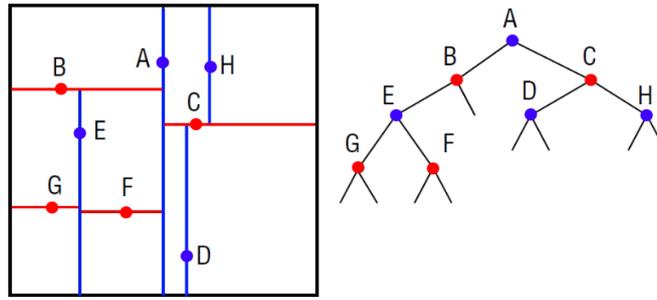


Figure 3.5: KD-tree data structure (modified from van Oosterom [1999]).

Where $f_{deviation}$ is the deviation of the feature value from the image mean, f_i is the feature value at segment i , \bar{f}_w is the global weighted mean of the feature (computed using all pixels in the image) (see Equation 3.12).

In addition to slope, another topographic feature named relative relief is computed from ALOS World 3D-30m DEM. It is calculated as the difference between the highest and lowest points in elevation within each segment Equation 3.14:

$$H_r = h_{max} - h_{min} \quad (3.14)$$

Where H_r is the relative relief, h is the corresponding elevation value at each pixel.

3.3 IMAGE CLASSIFICATION

Object-based classification is the process of assigning objects to classes based on spectral, spatial, contextual and textural properties (Section 2.2.2). In OBIA, classification can be performed applying different approaches [Bunting et al., 2014]:

- Rule-based approach: manually defined decision trees;
- Unsupervised Classification: clustering algorithms;
- Statistical Supervised: e.g. Maximum Likelihood Classification(MLC), Mahalanobis Distance, Parallelepiped Classifier;
- Supervised Machine Learning: e.g. Random Forest (RF), Support Vector Machines (SVM).

Using features extracted from Sentinel-2 imagery and ALOS World 3D-30m DEM, a supervised ML approach is applied to classify landslides candidates outlined through segmentation. The classification is binary: *landslide* or positive class (labeled as 1) and *non-landslide* or negative class (labeled as 0).

Previous scientific works [Melgani and Bruzzone, 2002; Stumpf and Kerle, 2011; Mountrakis et al., 2011; Puissant et al., 2014; O'Connell et al., 2015] highlight that non-parametric techniques such as SVM and RF are more suitable for object-based image classification of remotely sensed images than parametric ones as they can handle different statistical distributions. SVM demonstrates good performance over high dimensional data and with small training sets. Drawbacks of the algorithm are the appropriate selection of the kernel function and the complexity in tuning parameters.

RF classifier is found to be the most appropriate ML classification technique based on the following considerations: at an initial stage of this study, both classifiers

SVM and RF were used, with RF showing more promising results. In the context of OBIA, it has demonstrated high performance for object-based classification of optical satellite imagery [Lebourgeois et al., 2017; Ma et al., 2017; Stumpf and Kerle, 2011; Parker, 2013; Stumpf and Kerle, 2011; Pal, 2005] with just a few relevant tuning parameters including the number of trees, trees depth, and the number of features used per split. It can deal with a great variety of statistical distributions, performs an adequate generalization when data is missing, handles imbalanced datasets using BRF and WRF (see Section 2.5.2), and it can support the selection of features by measuring their contribution to the final decision of the model.

3.3.1 Training samples and labeling

A subset of the samples database created in Section 3.1.1 is linked to the target classes and subsequently used to train the model. Areas in the image containing landslides polygons are labeled as the *landslide* segments, while the remaining areas are labeled as *non-landslide*.

To generalize the model, landslides of different sizes and types occurring across various landscapes, and induced by different triggering factors (e.g. rainfall, earthquakes, snow avalanches) should be part of the training samples.

3.3.2 Random Forest Classifier

For an introduction of the early developments and definition of RF, the reader is referred to Section 2.3. The mathematical foundations and basic principles of the algorithm are addressed in this section.

Mathematical foundation

In ensemble classification, multiple classifiers are used to predict the most frequently occurring class. When the training dataset is input into a decision tree, it formulates a set of rules and conditions that will be used to make the predictions. The two main decision-tree algorithms used in RF to define those rules, evaluate the quality of a split, and iteratively divide the dataset into regions are:

1. Classification and Regression Tree (CART) [Breiman et al., 1984]. For CART the Gini index is used as a measure of node impurity (quality of a split) when deciding to split. This metric measures the divergences between the probability distributions of the target attributes value [Singh and Gupta, 2014].

Given a number of k classes in the dataset, the Gini index is defined as a measure of total variance across the k classes (Equation 3.15):

$$\mathcal{G} = \sum_{k=1}^k P_k(1 - P_k) \quad (3.15)$$

Where P_k is the proportion of the number of elements in class k . When P_k is close to 0 or 1, the Gini index results in a small value. The smaller the values of the Gini index, the higher the purity of the node; which indicates that the node predominantly contains observations from a single class [James et al., 2013].

2. Iterative Dichotomiser (ID₃) [Quinlan, 1986]. To construct a decision tree, ID₃ develops a classification rule that can determine the class of any object from its attribute values. The main idea of ID₃ is to generate all possible decision trees that correctly classify the training set and then select the simplest of them.

Entropy and Information Gain are the mathematical principles used to select the best split and evaluate the purity of the node [James et al., 2013].

The Entropy or the measure of the amount of uncertainty in the dataset is defined by Equation 3.16:

$$\mathcal{H}(S) = - \sum_{k=1}^k P_k \log P_k \quad (3.16)$$

Where P_k is proportion of the number of elements in class k to the number of elements in set S , and S is the current dataset.

Similar to the Gini index, the Entropy takes a minimum value when all observations in the dataset fall into a single target class (high node purity). The Information Gain is calculated using Equation 3.17:

$$\mathcal{G}ain(S, a) = H(S)_{t_1} - H(S|a)_{t_2} \quad (3.17)$$

Where Gain (S, a) is the Information Gain (decrease in entropy), H is the entropy, t_1 is the *prior* state, t_2 is the state after the dataset is split, and a is the attribute evaluated to split.

In general terms, ID3 chooses its splits based on the decrease in Entropy between the parent node (weighted sum) and the children's nodes. The main idea when building the decision tree is to find the attribute that returns the highest Information Gain.

Splitting and feature selection

The splitting criterion in a classification decision tree aims at decreasing the impurity of a node using metrics such as the Gini index (for CART) and Information Gain (for ID3) to evaluate the class mixing and the quality of a particular split. The splitting measures are defined in terms of the class distribution before and after splitting [Singh and Gupta, 2014].

The basic principles of RF to select in which feature to split on are:

- When a split in an individual tree is considered, a random sample of m predictors is chosen as split candidates from the full set of p predictors [James et al., 2013; Fawagreh et al., 2014];
- The number of m predictors considered at each split is commonly defined by

$$m = \sqrt{p} \quad (3.18)$$

Where p is the total number of predictors in the dataset, and m is a subset of p .

James et al. [2013] state that choosing the best split among a limited number of predictors avoids building highly correlated trees. When the p predictors are given as an option in a collection of trees, most of them will use the strongest predictor as the top split, leading to very similar trees with an average prediction with high variance and, therefore, less reliable. In RF, approximately $(p - m)/p$ of the splits will never consider the strong predictor, allowing other predictors to be selected as the root node [James et al., 2013].

Bootstrap sampling

In addition to the random selection of features, **RF** uses Bagging or **Bootstrap Aggregation** to build the trees. **Breiman** [1996] defines Bagging as a method for generating multiple versions of a predictor and using them to get an aggregated predictor. **Bootstrap Aggregation** is used to reduce the **variance** and increase the prediction accuracy of statistical learning methods [**James et al.**, 2013]. When creating an individual tree in **RF**, a new training set (bootstrapped training samples) is drawn with replacement from the original dataset [**Breiman**, 2001b].

Each bootstrap training set uses about one-third of the instances or observations. The remaining set of observations (not used to fit a given bagged tree) are the **Out-of-Bag (OOB)** observations [**Breiman**, 2001a]. At each iteration, the algorithm predicts **OOB** data using the tree growth with the bootstrap sample and calculates the classification error [**Biau and Scornet**, 2016]. Since the error rate decreases as the number of combinations increases, the **OOB** estimates will tend to overestimate the current error rate [**Breiman**, 2001b]. Given that the response for each observation is predicted using only the trees that were not fit using that observation, the resulting **OOB** error is a valid measure of the test error for the bagged model [**James et al.**, 2013].

In sum, using the principles of **Bootstrap Aggregation**, **RF** improves its performance by taking many training sets from the dataset, building a separate prediction model using each training set, and averaging the resulting predictions [**James et al.**, 2013].

Majority voting

To predict the class of unlabeled observations, **RF** aggregates individual trees predictions and uses a voting scheme to get a final decision. One of the simplest and most effective voting schemes is the majority voting proposed by **Lam and Suen** [1994]. In this method, each classifier is asked to predict the class label of the tested observation; the class that receives the greatest number of votes is returned as the final decision of the **ensemble** [**Fawagreh et al.**, 2014].

Feature importance

There are different methods implemented in **RF** to rank features [**Breiman**, 2002]; two of the most common ones are the Mean Decrease Impurity (**MDI**) and Mean Decrease Accuracy (**MDA**) [**Biau and Scornet**, 2016].

When using **MDI**, **RF** computes how much each feature decreases the weighted impurity in an individual tree [**Biau and Scornet**, 2016]. Using either the Gini index or Information Gain, the features are ranked by averaging the total decrease in node impurity over all trees in the forest.

In the case of the **MDA**, **RF** estimates the importance of an individual feature by looking at how much prediction error increases when the **OOB** for that feature is permuted while all others are left unchanged [**Liaw et al.**, 2002]. The process is repeated for every single tree in the forest and the **MDA** of the analyzed feature is obtained by averaging the difference in **OOB** error estimation before and after the permutation over all trees. **Biau and Scornet** [2016] state that **MDA** is based on the idea that if the feature is not relevant, then rearranging its values should not degrade the prediction accuracy [**Biau and Scornet**, 2016].

3.3.3 Training phase

A subset of the initial number of features extracted from Sentinel-2 and **DEM** (see **Table 3.1**) are used to train the **RF** classifier.

The approach follows in this study is first to focus on valuable features that *a priori* provide diagnostic information to discriminate between the *landslide* and the

non-landslide class. The classification strategy is iterative (Figure 1.1 c); that means that *landslides diagnostic features* are progressively included as training features, and the model performance is evaluated at each step. To assess the extent to which every of the input features is discriminant to build the model, the Mean Decrease Impurity (MDI) is used.

3.3.4 Model assessment

The model assessment is crucial to evaluate how well the RF classifier generalizes to unseen data. When a large dataset is not available, alternative approaches to conducting the accuracy assessment are resampling methods as the cross-validation technique that consists in estimating the test error rate (probability error) by leaving out a subset of the training samples before fitting the model and used it to test the classifier performance [Dougherty, 2012].

A useful and straightforward strategy for model evaluation is the validation set or holdout approach. It consists in splitting the dataset into two sub-groups: a training set and a validation or testing set. The classifier is fit on the training set, and the fitted model is used to make predictions for the observations in the test set [James et al., 2013].

For this study, the proportion of the data used for train and test the model is 70% and 30%. The model performance is evaluated using the confusion matrix and those metrics suitable for imbalanced datasets, including precision, recall, and f-measure (Section 2.5.3).

4

DATASETS AND TOOLS

4.1 DATASETS

4.1.1 Sentinel-2 mission

Sentinel-2 mission is selected as the input dataset as it is the free-of-charge optical satellite imagery with the highest Ground Sample Distance (GSD). The European Space Agency (ESA) launched the mission as part of the Copernicus program (Sentinel-2A: June 2015; Sentinel-2B: March 2017). It comprises a constellation of twin polar-orbiting satellites placed in the same orbit, phased at 180° to each other¹. Sentinel-2 aims at monitoring variations in land surface conditions, including vegetation, soil, and water cover. The constellation has a global spatial coverage between latitudes 56° south and 83° north.

The products comprise 13 spectral bands with variables GSD representing Top-Of-Atmosphere (TOA) reflectance scaled by 10,000 (Table 4.1). The bands span the visible, Near-Infrared (NIR) and Short-Wave Infrared (SWIR) of the electromagnetic spectrum. Additional QA60 bands are included to support the detection and removal of clouds.

Band	GSD (m)	Central wavelength (nm)	Description
B1	60	443	Aerosols
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Red Edge 1
B6	20	740	Red Edge 2
B7	20	783	Red Edge 3 ¹
B8	10	842	NIR
B8A	20	865	Red Edge 4
B9	60	945	Water vapor
B10	60	1375	Cirrus
B11	20	1610	SWIR ₁
B12	20	2190	SWIR ₂
QA60	60	-	Cloud mask

Table 4.1: Sentinel-2 spectral and QA bands.

For this study, Level-1C products (Sentinel-2 Multispectral Instrument Level-1C, TOA) are used. They have radiometric and geometric corrections, including orthorectification and georeferencing on a global reference system (WGS84) with sub-pixel accuracy. The products are resampled with a uniform GSD of 10m, 20m and 60m depending on the resolution of the different spectral bands [SUHET, 2013]. More details of the Sentinel-2 mission are given in Table 4.2.

The detection of clouds conducted by Sentinel-2 providers using the QA60 bands is based on the following considerations [ESA, 2013]:

- Dense or opaque clouds are characterized by high reflectance in the blue spectral region;

¹ http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview4

Revisit	5 days at the Equator
Number of bands	13 spectral bands
GSD	10m, 20m or 60m
Temporal coverage	2015-Present
Spatial coverage	56°S - 84°N
Swath width	290km
Orbit	Polar-orbit, sun-synchronous
Mean orbital altitude	786km
Tiles size (level-1C)	100km x 100km orthoimages
Projection (level-1C)	UTM/WGS 84
Radiometric resolution	12 bits

Table 4.2: Sentinel-2 mission details.

- As the snow also has a high reflectance in the Blue band (B₂), SWIR in B₁₁ and B₁₂ are used to identify clouds. In contrast to snow, clouds have a high reflectance in SWIR;
- As high altitude ice clouds have low reflectance in the SWIR (B₁₁ and B₁₂), the high atmospheric absorption band (B₁₀) is used to avoid confusion with snow;
- Cirrus or semi-transparent clouds can not be identified in the B₂. Pixels with low reflectance in the B₂ and high reflectance in the B₁₀ has a high probability of being cirrus cloud.

4.1.2 Digital Elevation Model

Digital Elevation model (DEM) AW_{3D30} is selected as the elevation input dataset due its GSD (30m) and its acquisition time. It is the most updated (2006-2011) free-of-charge DEM with global coverage (Table 4.3).

AW_{3D30} derived from the Japan Aerospace Exploration Agency (JAXA) that uses the archived data of the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM) onboard the Advanced Land Observing Satellite (ALOS). The dataset contains a horizontal GSD of approximately 30m mesh (1 arc-seconds latitude and longitude) generated from the 5m resolution Digital Surface Model (DSM) [Tadono et al., 2014]. Using an updated version of the Delta Surface Fill method (March 2017), the void height values in cloud and snow pixels between 60° south and 60° north latitudes are filled with existing DEMs. This method consists in replacing the void values in the original DEM with the adjusted values calculated from surrounding valid pixels in other references DSMs [JAXA, 2017].

GSD	30m
Acquisition time	2006-2011
Spatial coverage	60°S - 60°N
Tile size	1° x 1°
Projection	UTM/WGS 84
Minimum value	-479
Maximum value	8859

Table 4.3: DEM ALOS description.

4.1.3 Software specifications

GOOGLE EARTH ENGINE (GEE) is used to pre-processes the dataset. GEE is a cloud-based platform and free JavaScript programmable interface that offers global

coverage of earth-observation data. It combines a multi-petabyte catalog of remotely sensed and geospatial datasets with the capabilities of planetary-scale analysis. Its parallel processing capacity makes it efficient to run spatial reductions over large image collections [Gorelick et al., 2017], supporting the process of handling space-time analysis for image set preparation pre-and post-events on a global scale. **GEE** library counts with more than 800 functions, which range in complexity from simple mathematical functions to powerful geostatistics and image processing operations [Gorelick et al., 2017]. Additionally, available functions as `map()`, `iterate()` allow applying a wide variety of functions to collections of images in parallel.

PYTHON is the main programming language used for the processing and visualization of the data. Python-Based Open Source System for Geographic Object-Based Image Analysis (**GEOBIA**) is used for the initial segmentation and generation of attribute tables (features statistics per segments) [Clewley et al., 2014]. The system uses the following set of Python packages:

- **Numpy** for raster data processing;
- Remote Sensing and GIS library (**RSGISLib**) for segmentation and attribution of objects;
- Geospatial Data Abstraction Library (**GDAL**) for geospatial data manipulation and raster data model;
- The Raster Input and Output Simplification (**RIOS**) for reading and writing raster attribute tables, which are represented internally as NumPy arrays;
- Tuiview for viewing and manipulating raster attribute tables;
- Kea for storing image objects and associated attributes.

Scikit-learn, a free software **ML** library is used for the classification stage. **Seaborn**, and **Matplotlib** are used for visualization.

POSTGRESQL is used as the database management system for storing the landslide inventory tables described in [Section 5.1.1](#). It allows data handling among several datasets and multiple non-spatial and spatial operations using the PostGIS extension.

QGIS is used as the tool for creating the landslides polygons and for editing and visualizing the datasets tables stored in PostgreSQL.

5

EXPERIMENTS AND RESULTS

This chapter describes the implementation details of the methodology and presents the experiments and the results of this study. The source code of this implementation can be found at https://github.com/mhscience/landslides_detection. The chapter is structured as follows: [Section 5.1](#) gives details of the built landslide database. [Section 5.2](#) describes different strategies and parameters used to produce image segments and balance the dataset. Finally, [Section 5.3](#) gives an overview of the experiments and tuning hyperparameters used to train the RF model and presents the final results of the model assessment. Analysis of the results is presented in [Chapter 6](#).

5.1 PRE-PROCESSING

5.1.1 Sample set preparation

Using pre- and post-event images within Google Earth Engine ([GEE](#)), we recognize 110 landslides in 32 satellites images distributed worldwide ([Figure 5.1](#)) ([Section 3.1.1](#)). We digitize and store the landslides polygons as shapefiles that are later used for class labeling in the classification stage. The samples are stored in PostgreSQL as tables containing the landslide/landslides identification (*landslide_id*), geographical coordinates (latitude, longitude), event date, triggering factor, landslide size ([Table 5.1](#)), area/length/width of the major landslide recognized in the image, number of landslides occurrences per image, the source of the information, land cover class, the quality assessment of the samples, and extra information described in [Appendix A](#). A summarized representation of the landslide inventory table is depicted in [Table 5.2](#).

Sizes	Area (m ²)
Medium	> 250 and ≤ 1,000
Large	> 1,000 and ≤ 100,000
Very large	> 100,000 and ≤ 500,000
Catastrophic	> 500,000

Table 5.1: Landslide sizes.

We made a visual and coarse manual labeling per satellite image (land cover column) that contains the following classes: urban, vegetated areas (e.g. forest, grasslands, shrublands), barelands, and croplands. A mixture of classes is commonly identified per image; when a sample is not labeled as urban, then it is assumed that the landslide or landslides events occurred in a remote area ([Table 5.2](#)).

We perform a quality assessment study (quality column) that consists in a qualitative evaluation per image to determine the degree of confidence in using the samples as part of the training set. This assessment is based on the quality matrix illustrated in [Table 5.3](#). The rows of the matrix are conditioned by the geographical location (GI), the landslide size, and the land cover. A sample with a validated geographical location has a higher quality than others without validation. The landslide sizes consider to make the assessment are catastrophic (high quality), very



Figure 5.1: Global coverage of the sample set. The red dots represent the location of the samples images. Given the large-scale of the map, some dots are overlapped in specific areas such as Japan and New Zealand.

id	event_date	longitude	latitude	location_accuracy	landslide_trigge	size	country	# events	quality	land cover	Source
L0	2017-03-25	-76.66247976	1.169677465	exact	downpour	very large	Colombia	1	M1	Vegetated areas/ Urban	NASA Catalog
L1	2017-06-27	103.6529857	32.06849057	exact	continuous_rain	catastrophic	China	1	H1	Vegetated	NASA Catalog
L2	2017-08-14	-13.22985664	8.436115153	exact	unknown	very large	Sierra Leone	1	M1	Vegetated areas /Urban	Web resources
L3	2017-01-10	-65.46772261	-23.9134425	approximated	downpour	very large	Argentina	1	S1	Bare-lands	NASA Catalog
L4	2016-05-18	80.4319754	7.16080832	exact	monsoon	very large	Sri Lanka	2	M1	Vegetated areas	NASA Catalog
L5	2016-06-15	-117.2679	56.2246	approximated	downpour	large	Canada	1	S3	Croplands	NASA Catalog
L7	2016-11-27	101.613538	3.366921	exact	rain	large	Malasia	1	M2	Urban	NASA Catalog
L12	2017-02-09	-43.42657247	-20.24799736	exact	mining	very large	Brasil	1	H3	Wetlands	NASA Catalog
L17	2017-01-18	13.77821597	42.42921449	exact	earthquake	large	Italy	2	H3	Vegetated area	NASA Catalog
L25	2017-07-06	130.8207411	33.40312701	exact (major landslide)	tropical_cyclone	various	Japan	3	M2	Vegetated areas	Time series GEE
L41	2018-04-11	130.8966053	33.43075427	exact	unknown	large	Japan	1	H3	Vegetated areas	Web Resources
L55	2016-11-15	173.8168708	-42.20224732	bbox center	earthquake	large	New Zealand	2	H3	Vegetated areas	Web Resources

Table 5.2: Landslide database example. The location accuracy "bbox center" means that the geographical coordinates represent the center of the bounding box used to download the image (see [Figure 3.2](#)). For an extended version of the database see [Appendix A](#).

large (medium quality) and large (low quality). Respect to the land cover classes, a satellite image entirely represented by vegetated areas has a higher quality than others with a mix of vegetated-urban; the lowest quality is assigned to samples represented by barelands with sparse of non-vegetation.

The columns of the matrix are the cloud (C) and snow percentage (S) that remain after pre-processing ([Section 3.1.2](#)). A sample has higher quality while lower the could or snow percentage is. Rows and columns of the matrix are combined to give a final quality assessment per sample. As indicated in [Table 5.3](#), we define a total number of nine classes; the quality of a sample progressively increases from label S3 (lowest quality) to label H1 (highest quality). The assessment conducted to the entire dataset ([Appendix A](#)) resulted in 16 samples labeled with high quality (2:H1, 3:H2, 11:H3) 16 with medium quality (8:M1, 6:M2, 2:M3), and four with low quality (1:S1, 3:S3). Samples labeled with lower quality than M3 are not used to train/test the model.

5.1.2 Image set preparation

Exploiting the capabilities of [GEE](#) for efficiently handling big earth observation datasets, we conduct the image set preparation within this cloud-based platform. The implemented algorithm ([Algorithm 3.1](#)) finds the less clouded images within a pre-determined period, creates cloud-free composites, and removes clouds using cloud masking. [Table 5.4](#) presents a summary of the sample and image set inven-

Description	C or S <15%	15% < (C or S) <40%	Scale
GL validated / catastrophic landslides/vegetated areas	H1	H3	H1
GL validated / very large landslides/vegetated areas	H2	M1	H2
GL validated / large landslides/vegetated areas	H3	M2	H3
GL validated / catastrophic landslides/ wetlands or croplands	H2	M1	M1
GL validated / very large landslides/wetlands or croplands	H3	M2	M2
GL validated / large landslides/wetlands or croplands	M1	M3	M3
GL validated / catastrophic landslides/vegetated areas and urban	H3	M2	S1
GL validated / very large landslides/vegetated areas and urban	M1	M3	S2
GL validated / large landslides/vegetated areas and urban	M2	S1	S3
GL validated /various size/barelands	S2	S2	
Uncertain Geographical Location + any of the above options	S3	S3	

Table 5.3: Quality assessment matrix for the samples set. GL is the geographical location of the major landslide recognized in the image, C is the cloud percentage, and S is the snow percentage.

tory. The effect of image composition and cloud masking in pre- and post-event images is illustrated in [Figure 5.2](#).

Description	Quantity	Remark
Retrieved images	96	post- and pre-event/image difference
Processed images (GEE)	> 1,500	≈ 1 TB
Landslides	110	heads-up digitization
Land cover	4	urban/vegetated/barelands/croplands
Quality assesment	9	H1/H2/H3/M1/M2/M3/S1/S2/S3

Table 5.4: Summary of the image and sample sets inventory.

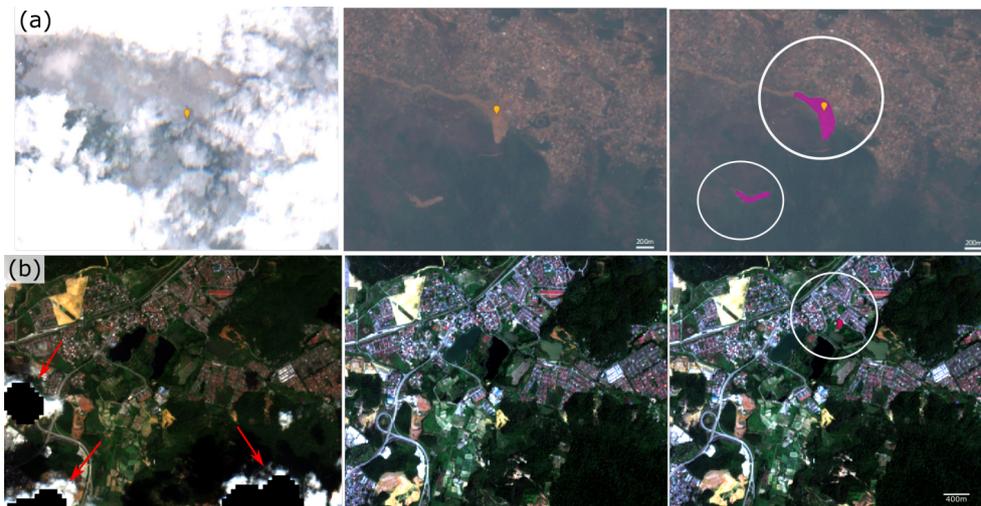


Figure 5.2: Generation of cloud-free images pre- and post landslide event and heads-up digitization of landslides polygons. (a) Cloud-free images are created using the image composition technique explained in [Section 3.1.2](#) (Landslide L2). (b) Clouds are removed using cloud masking (Landslide L7).

Parameters	Final values	Description
numClusters	19	Number of clusters
minPxls	80	Minimum number pixels within a segments
distThres	100	Distance threshold for joining the segments
bands	<i>rgd</i>	Sub-set of image bands to use
kmMaxIter	200	Maximum iterations for k-Means

Table 5.5: Initial segmentation parameters.

5.2 IMAGE SEGMENTATION

5.2.1 Initial segmentation

We perform the initial image segmentation in Python `RSGISLib` using the [Shepherd et al. \[2019\]](#) algorithm (Section 3.2.1). For this study, the optimal segmentation scale is defined as the scale at which landslides candidates of different spectral signatures and sizes can be identified as non-mixed segments. This is achieved by using as input segmentation feature the Red/Green Difference (RGD) (see Equation 3.5) and experimentally tuning a set of parameters that are illustrated in Table 5.5.

To compare segmentation results derived from different input features, the following experiments are conducted:

1. Segmentation on image difference using RGD.
2. Segmentation on post-event images using absolute NDVI values (see Equation 3.2);
3. Segmentation on post-event using Red band (B_4), Green band (B_3), and B_2 ;

Results demonstrated that RGD is the feature that better highlights the changes induced by landslides and reduces variations caused by other irrelevant factors (e.g. changes in illumination). Using absolute values of NDVI mostly fail in cases where large-scale landslides occur within urban environments or when they are contaminated with vegetation (Figure 5.3). The segmentation conducted using the three spectral bands B_4 , B_3 , B_2 does not work well for most of the samples. It generates a lower number of *non-landslide* segments but induces the class mixture.

The key parameters that condition the quality of the segmentation in terms of its capability to preserve landslides candidate segments (avoiding class mixture) are the minimum number of pixels and the number of clusters k . The minimum number of pixels is set to $80px$; a value slightly lower than the smallest size of the target landslides defined within the scope of this study (Section 1.4). The optimal number of clusters is first estimated by applying the Elbow method (Section 3.2.1). In Appendix B four plots examples of the reduction of the sum of squared errors (WSS) with the increase in the number clusters k are displayed. The plots show comparable results, where the points of inflection (elbows) or optimal number of clusters approximated in the same value. As the same behavior is achieved by running the k-means clustering over all the training images, a unique minimum k number is initially estimated in a value of eight for all images.

This k value is subsequently increased and experimentally adjusted to reduce undersegmentation of large-scale landslides (Figure 5.4). Table 5.5 shows the final value of k set to 19; this value increases the number of segments from 54,696 ($k=8$) to 55,470 ($k=19$) but reduces class mixture errors within the same segments.

5.2.2 Conversion of features per pixel to features per segments

Once the primitive segments are created, features per pixel are grouped to produce features per segment. We calculate the features statistics within segments polygons

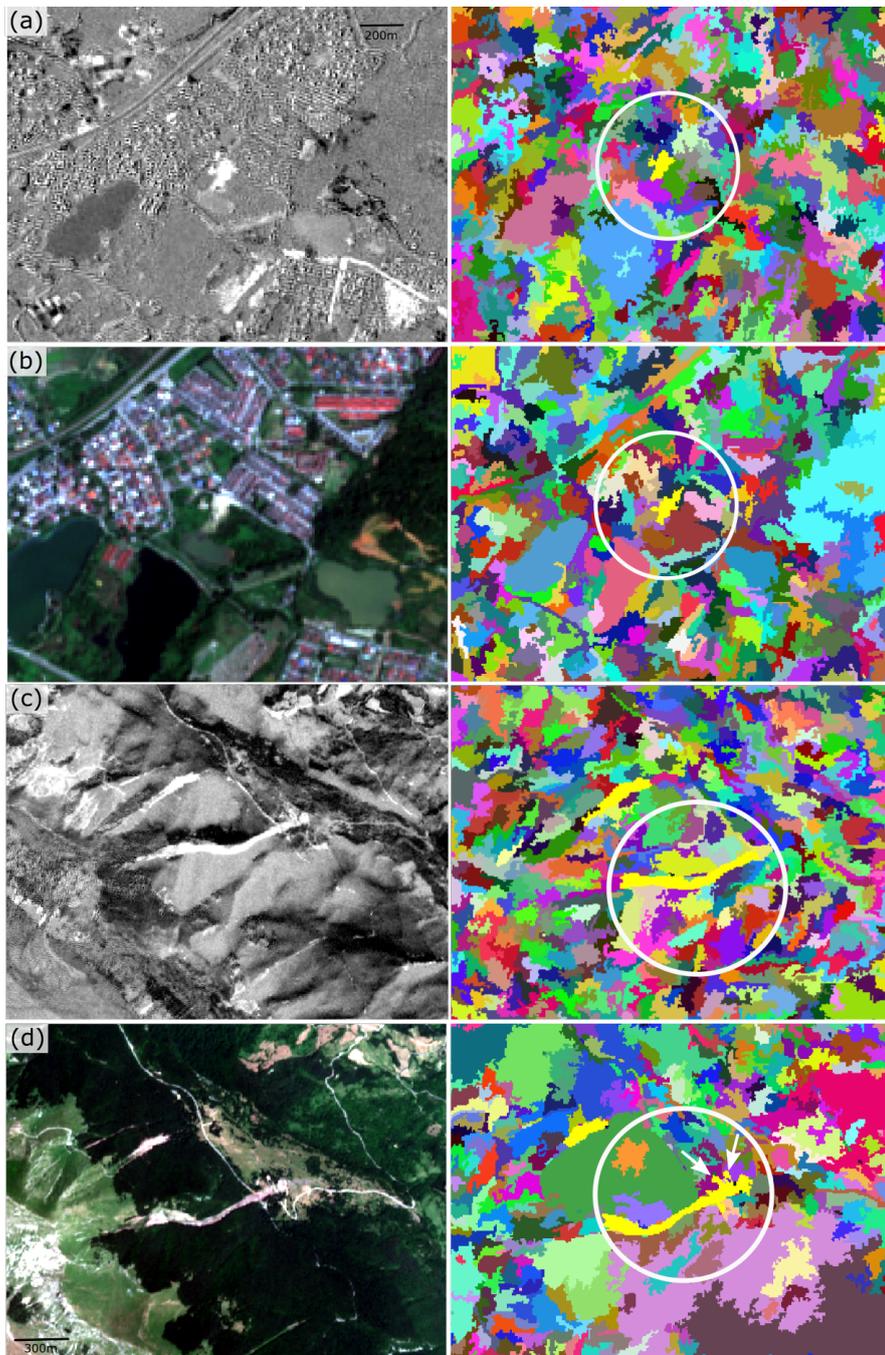


Figure 5.3: Comparison of segmentation results using different input features. (a) Segmentation on image difference using RGD (landslide *L7*); landslide boundaries well defined within an urban area (b) Segmentation on post-event image using NDVI; landslide and building pixels are merged into the same segment. (c) Segmentation on image difference using RGD (landslide *L17*); landslide boundaries well defined in a remote area. (d) Segmentation on post-event image using bands B_4 , B_4 , B_2 ; landslide boundaries (white arrows) are merged to a small deforested area.

using **RSGISLib** (Figure 5.5) and generate attribute tables per image with rows representing the segments and columns the values of the features (Table 5.6). As Table 5.7 indicates, the statistical measure calculated for all **landslides diagnostic features** is the mean. For the slope, the maximum values per segment is also calculated.

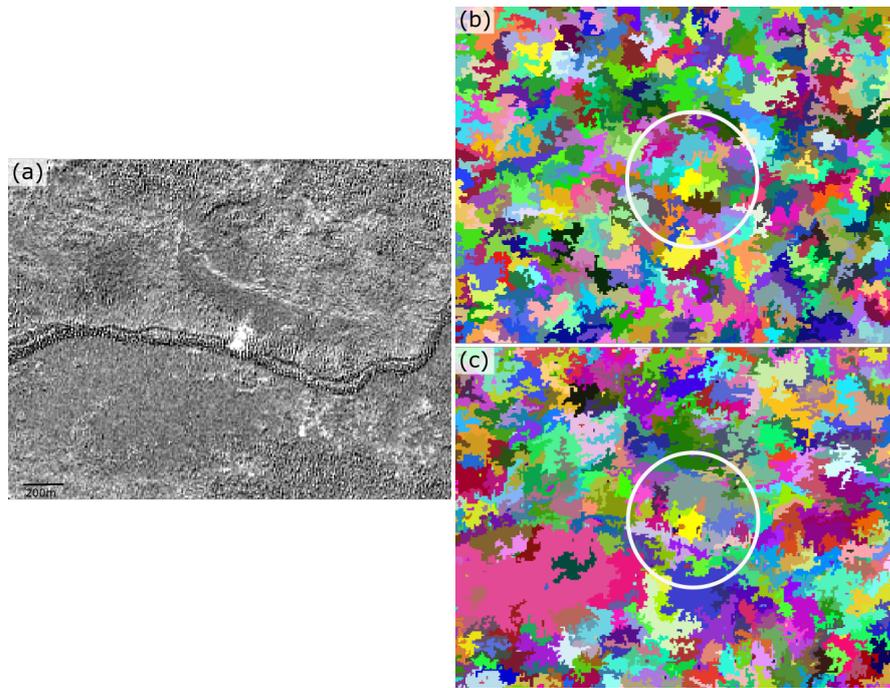


Figure 5.4: Initial segmentation using two values of k ($L9$). (a) Image difference using RGD. (b) Segmented image, $k=19$; large-scale landslide is correctly segmented. (c) Segmented image, $k=8$, the *landslide* is undersegmented and merged with *non-landslides*.

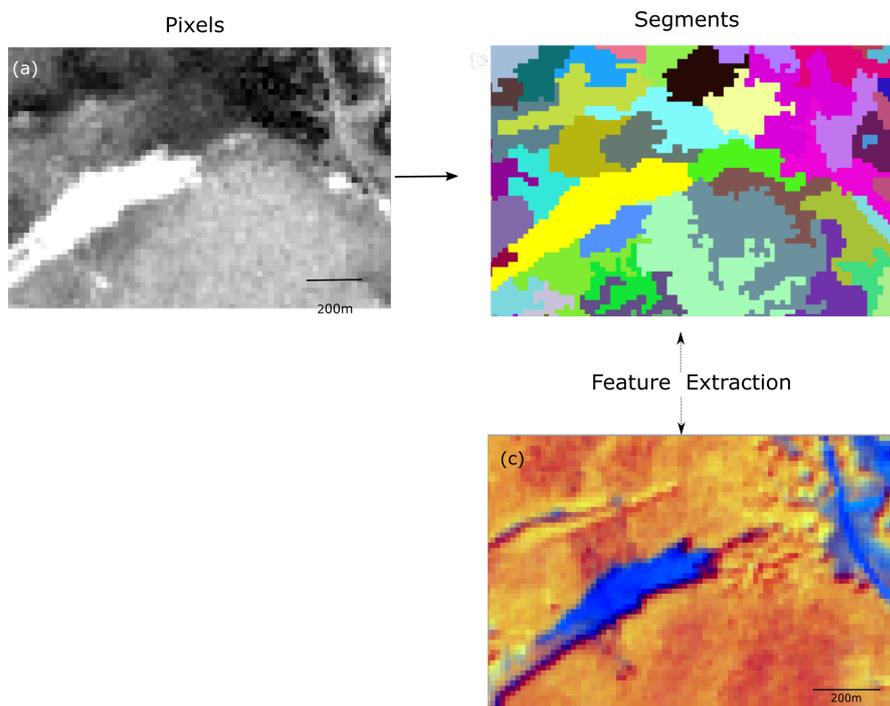


Figure 5.5: Pixels to Segments. Features statistic at segment level. (a) Image Difference (b) Segmented image. (c) Post-event image.

5.2.3 Merging algorithm

We develop the merging algorithm based on region growing within Python ([Algorithm 3.3](#)).

	NDVI	RGD	VID	BrightnessD	Slope_max	Slope_mean	Class	Class_name
518	0.23	0.66	0.09	-0.24	20.07	16.12	0	no_landslide
519	0.64	0.60	0.13	-0.23	33.86	23.62	0	no_landslide
520	0.51	0.54	0.09	-0.24	7.21	5.11	0	no_landslide
521	0.67	0.44	0.03	-0.09	38.45	33.62	0	no_landslide
522	0.54	0.62	0.17	-0.28	8.73	5.96	0	no_landslide
523	0.68	0.37	-0.08	-0.09	5.62	5.62	0	no_landslide
524	0.61	0.39	0.01	-0.03	7.44	4.85	0	no_landslide
525	0.61	0.49	0.03	-0.14	16.55	9.48	0	no_landslide
526	0.66	0.44	0.00	-0.23	11.14	7.94	0	no_landslide
527	0.20	0.85	0.49	-1.13	20.28	6.31	1	landslide
528	0.56	0.30	-0.05	0.05	7.44	5.97	0	no_landslide
529	0.60	0.54	0.06	-0.15	27.11	22.16	0	no_landslide
530	0.67	0.56	0.06	-0.20	4.98	2.55	0	no_landslide
531	0.67	0.42	-0.01	-0.12	16.15	7.37	0	no_landslide
532	0.31	0.81	0.30	-0.47	16.62	13.85	0	no_landslide
533	0.61	0.48	0.04	-0.15	33.58	15.74	0	no_landslide
534	0.60	0.51	0.05	-0.20	11.89	9.74	0	no_landslide
535	0.37	0.39	0.23	0.06	7.03	3.44	0	no_landslide
536	0.63	0.52	0.04	-0.21	31.36	13.93	0	no_landslide
537	0.61	0.36	-0.15	-0.07	4.98	4.98	0	no_landslide

Table 5.6: Attribute table example (L17).

Feature nature	Feature	Statistics per segment
Spectral	NDVI	mean
Spectral	GNDVI	mean
Spectral	Brightness	mean
Spectral	RGD	mean
Spectral	VID	mean
Spectral	BrightnessD	mean
Textural	NDVI _{texture}	mean
Spatial	Slope	mean, maximum
Spatial	Relative relief	mean
Contextual	RGD _{deviation}	mean
Contextual	VID _{deviation}	mean
Contextual	BrightnessD _{deviation}	mean

Table 5.7: Landslide diagnostic features at segment level.

The statistical distributions of **spectral features** vary considerably from one image to another; examples are the standardized features distributions of *L0*, *L2*, and *L16* depicted in Figure 5.7. Image samples exhibit different distributions: Gaussian, skewed left/right, or bi-modal; however, landslides tend to remain around the same locations of the distributions. As seen in Figure 5.7, NDVI distributions of the three sample images (Figure 5.6) show that landslides are located to the left of the image means, while in the case of **change features** such as Vegetation index difference (VID) (Equation 3.6) and Red/Green Difference RGD (Equation 3.5), they tend to be found on the right tail of the distributions (see Section 3.2.2).

To explore how to merge similar segments while preserving landslides, four experiments are conducted starting with a fixed **coarse-grained** NDVI similarity threshold (Equation 3.11) (0.1, 0.2, 0.3, and 0.4), and subsequently considering image **non-local** NDVI, VID, and RGD, and finally a **fine-grained** threshold of 0.05. The input is a sample of 32 primary segmented images that after outliers removal generated 41,529 *non-landslide* and 199 *landslide* segments. The outputs of these experiments are the percentage of segment reduction and the percentage error caused by wrongly merged *landslides*.

The first experiment explores a simple approach just merging segments with NDVI thresholds of 0.1, 0.2, and 0.3. Results in Table 5.8 indicate a continuous increase in the merging error as the threshold becomes **coarse-grained**. Adding the NDVI distribution and merging segments exclusively with values higher than the mean (Table 5.9) reduces the mislabeling error, but only merges around 50% of the segments. When VID is added, it is possible to merge over 80% of the segments with an

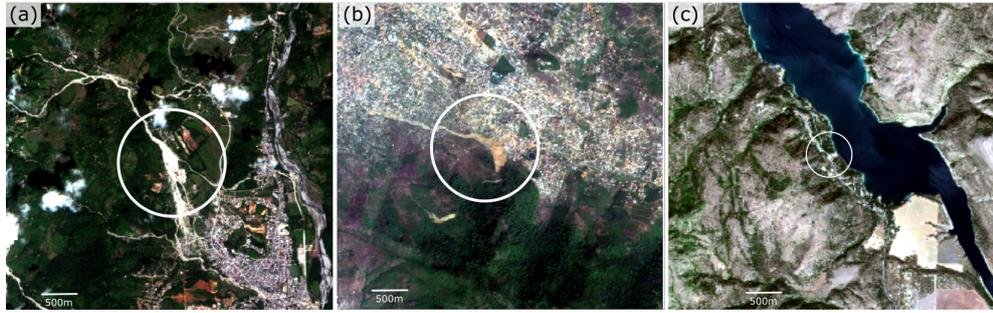


Figure 5.6: Post-event image of landslides L_0 , L_2 , and L_{16} . The white circle depicts the landslide location.

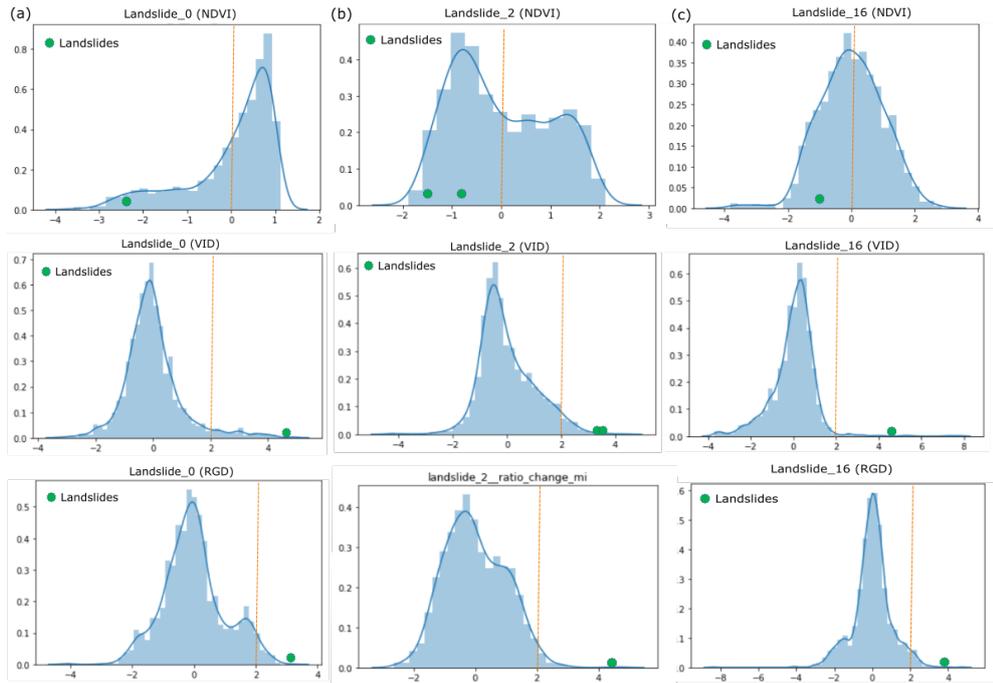


Figure 5.7: Standardized NDVI, VID, and RGD distributions for Landslides L_0 (a), L_2 (b), and L_{16} (c).

acceptable error of about 10%. *RGD* pushed the merging over 90% while maintaining the error at 8%. Finally, a *fine-grained* threshold with a value of 0.05, reduced approximately 100 more segments.

Coarse $Thrs_{NDVI}$	Final segments	Reduction	Missed	Error
0.1	7,132	82%	112	60%
0.2	2,407	94%	138	73%
0.3	1,059	97%	162	86%

Table 5.8: Merging algorithm using fixed NDVI thresholds.

To be conservative and keep the merging error below 10% and a high reduction of segments, the optimal *coarse-grained* threshold selected is 0.2. The use of a *fine-grained* threshold as the last merging strategy is useful to merge similar segments that might be closer to landslides in *NDVI*.

Coarse $Thrs_{NDVI}$	Final segments	Reduction	Missed	Error
Add NDVI				
0.1	20,095	56%	0	0%
0.2	18,929	54%	0	0%
0.3	18,671	53%	0	0%
Add VID				
0.1	7,779	81%	11	8%
0.2	3,342	91%	15	9%
0.3	2,007	95%	56	10%
Add RGD				
0.1	7,397	86%	11	5%
0.2	2,838	93%	16	8%
0.3	1,710	96%	24	13%
Add fine-grained threshold (0.05)				
0.1	7,296	82%	11	5%
0.2	2,749	93.3%	16	8%
0.3	1,653	96.3%	24	13%

Table 5.9: Merging algorithm using the distributions of spectral features. A sequential merging criteria is added: NDVI, VID, RGD, and a fine-grained threshold.

5.3 IMAGE CLASSIFICATION

5.3.1 Exploratory analysis

Once the attribute tables are derived from the merging algorithm, we conduct an exploratory analysis of the data in order to analyze the class distribution and identify discriminant and highly correlated features (Figure 5.8, Figure 5.9, and Figure C.1).

Redundant features such as the spectral bands (B_4 , B_3 , B_2) which are highly correlated with Brightness are not included in the model (Figure 5.8). Outliers and null-values that could decrease the model performance are filtered out based on features thresholds including low Brightness to remove shadows (<0.50), high Brightness to remove clouds (>2), and negative NDVI to remove water (<-0.1).

5.3.2 Training and testing

We conduct the training and testing phase using the RandomForestClassifier package available in the Python Scikit-learn library.

Input data

A total number of 2,905 segments derived from 32 satellite images (Appendix A) are used to train (70%) and test (30%) the RF model. From that total sample size, the *landslide* or positive class comprises 199 segments derived from 110 individual landslide events, while the *non-landslide* or negative class includes 2,706 segments.

Features selection

Using default tuning hyperparameters implemented in RF, the model is iteratively trained and tested by adding *landslides diagnostic features* and measuring their performance at each step (Figure 1.1, Section 3.3.3). An initial ranking of features (Table 5.10) that determines the order of addition is made based on relevant features used in previous researches (e.g. NDVI), and the features distribution derived from the exploratory analysis (discriminant features). As seen in Figure 5.9, *contextual* features such as $RGD_{deviation}$ and $BrightnessD_{deviation}$ (Section 3.2.3, Equation 3.13) are more appropriate than absolute values of RGD and Brightness to discriminate

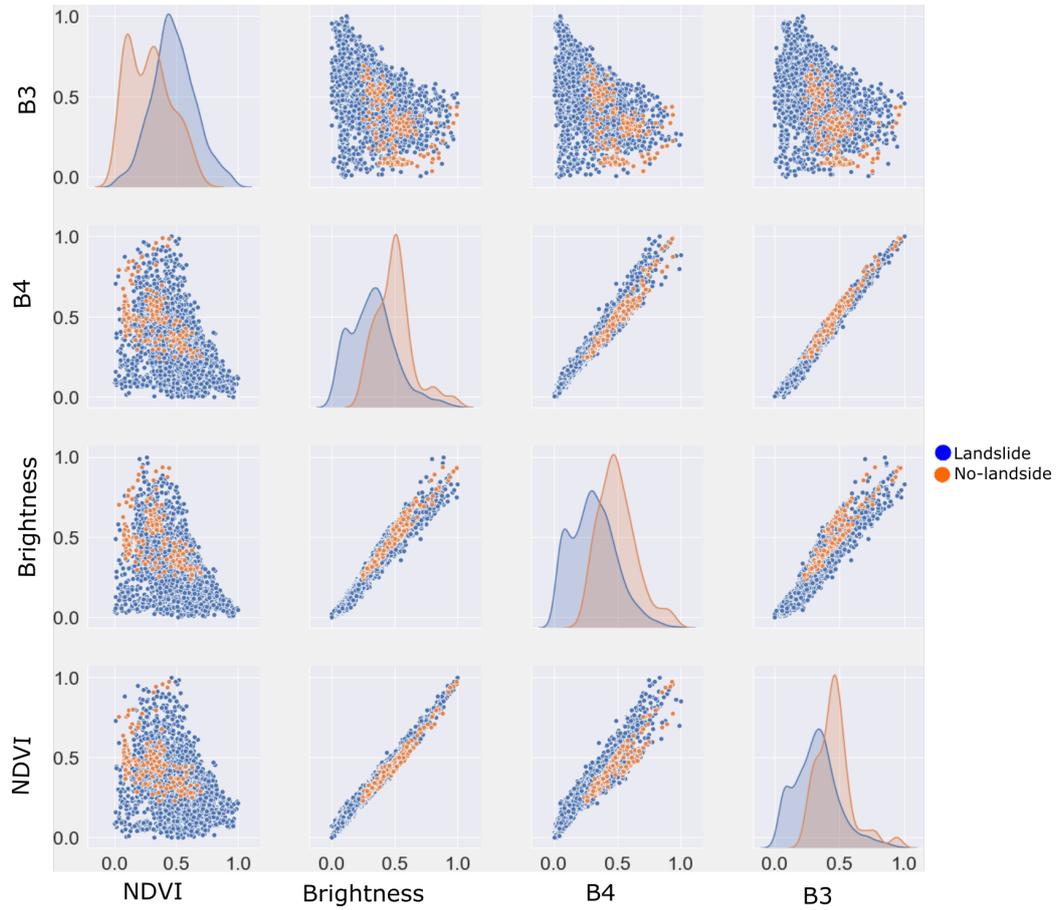


Figure 5.8: Class distribution and correlations between features: NDVI, Brightness, bands B4, and B2.

between *landslide* and *non-landslide* segments. Regarding the topographic features derived from DEM (Appendix C), it can be noticed that the distribution of the Slope_mean depicts a better separation between classes than Slope_max and Relative_relief.

Ranking	Feature
1	NDVI
2	RGD _{deviation}
3	BrightnessD _{deviation}
4	VID _{deviation}
5	Brightness
6	Slope_mean
7	GNDVI _{deviation}
8	Slope_max
9	NDVI _{texture}
10	Relative_relief

Table 5.10: Initial ranking of landslides diagnostic features.

Figure 5.10 shows the increase in model performance against the inclusion of new features (Table 5.10). The x-axis represents the features that are sequentially added, while the y-axis is the model performance in percentage.

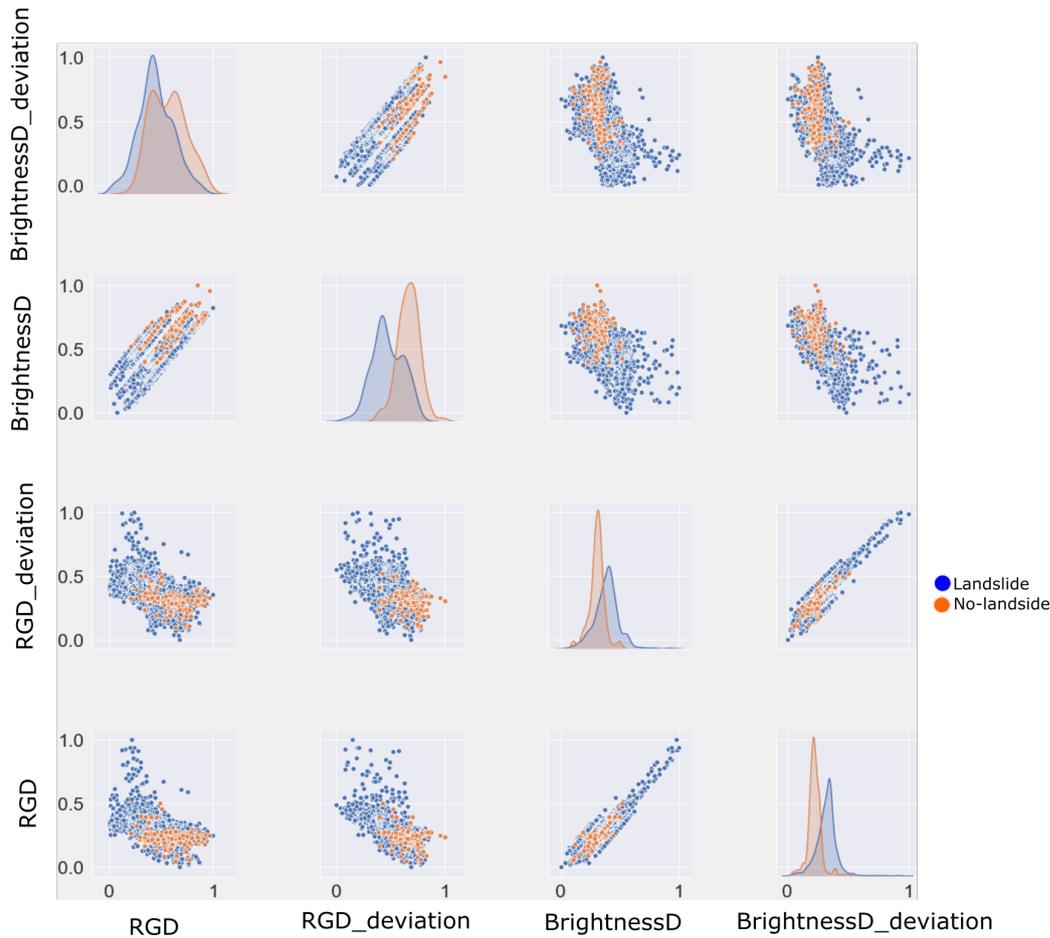


Figure 5.9: Class distribution and correlations between features: Red/Green difference (RGD), $RGD_{deviation}$, Brightness difference (BrightnessD), and $BrightnessD_{deviation}$.

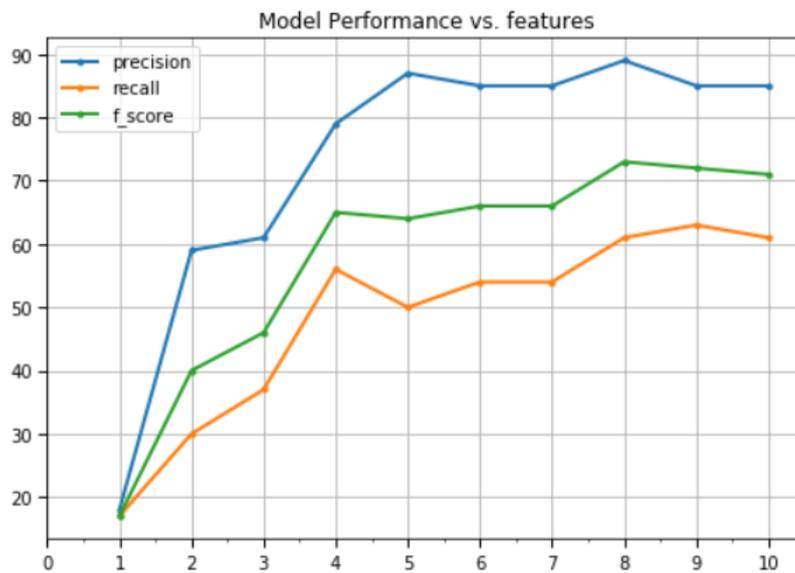


Figure 5.10: Model performance vs. features. Features represented in Table 5.10 are sequentially input to the model (1-10).

Random Forest classification

A set of RF hyperparameters are tuned in order to get the model with best predictive performance. The optimal values are experimentally obtained by testing all possi-

ble combinations using the GridSearchCV instance available in [Scikit-learn](#). The following values are tested:

- `n_estimators` or number of decision trees. It is tested for 10, 20, 40, 60, 80, 100, 150, 200, 250, 300, and 400;
- `max_depth` or maximum depth of the tree. It is tested for 5, 10, 20, 30, 35, 40, 45, 50, 60, and 70;
- `min_samples_split` or minimum number of samples required to split an internal node. It is tested for 2, 3, 4, 6;
- `min_samples_leaf` or the minimum number of samples required to be at a leaf node. It is tested for 2, 3, 4, 6;
- `max_features` or number of features to consider by each tree when splitting a node. It is set to "sqrt" to use a random selection of \sqrt{p} predictors ([Section 3.3.2](#));
- `class_weight` or weights associated with the classes. It is tested for 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:8, 1:9 and, 1:10 with higher weighted given to the minority class (*landslide* class);
- Bootstrap. It is set to True to use bootstrap samples when building trees ([Section 3.3.2](#)).

The best predictive performance of the model is achieved using the hyperparameters depicted in [Table 5.11](#).

Tuning hyperparameters	Final values
<code>n_estimators</code>	50
<code>max_depth</code>	40
<code>min_samples_split</code>	6
<code>min_samples_leaf</code>	4
<code>max_features</code>	sqrt
<code>class_weight</code>	1:5
<code>bootstrap</code>	True

Table 5.11: Optimal RF hyperparameters.

RF internally uses random samples with replacement (`Bootstrap = True`) to get a predictive model with less [variance](#) and therefore more reliable. Furthermore, the trees correlation is reduced by selecting a random subset of m at each split of the tree. The voting scheme adopted was the majority voting which is a suitable approach for classification.

With all the [landslides diagnostic features](#) shown in [Table 5.10](#) and the optimal hyperparameters, the model achieved promising results. The confusion matrix and the final model performance metrics for the *landslide* class (precision, [Equation 2.2](#); recall, [Equation 2.3](#); f-measure, [Equation 2.4](#)) are depicted in [Figure 5.11](#) and [Table 5.12](#). The f-measure is calculated using a β value of 1 to assign equal weight to both recall and precision (f1-score)([Equation 2.4](#))

Metric	(%)
Precision	83
Recall	83
f1_score	83

Table 5.12: Final model performance.

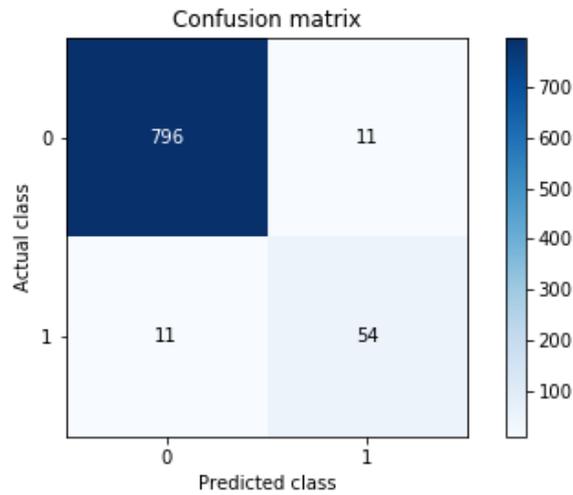


Figure 5.11: Confusion matrix (*landslides=1, non-landslides=0*).

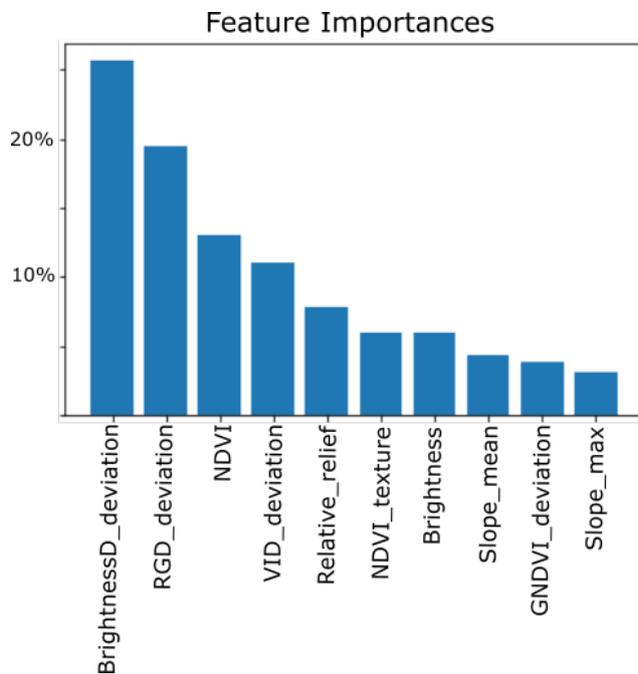


Figure 5.12: Final ranking of the landslides diagnostic features using Mean Decrease Impurity of RF.

A final assessment of the feature importance is made using the Mean Decrease Impurity (MDI) implemented in RF (see Section 2.5.2). Figure 5.12 illustrates the final ranking of the features by order of importance.

To evaluate the impact of using training samples labeled as urban or mixed urban/vegetated (see Section 3.1.2 and Appendix A), the model performance is assessed removing those samples (*L0, L2, L7, L9*) and keeping the same features and tuning parameters depicted in Table 5.11 and Table 5.10. The confusion matrix and metrics of this model are depicted in Figure 5.13 and Table 5.13.

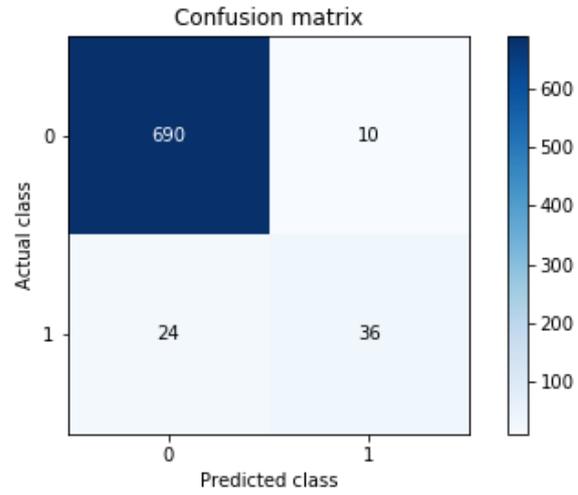


Figure 5.13: Confusion matrix excluding samples with mixed land cover: urban/vegetated (*landslides=1, non-landslides=0*).

Metric	(%)
Precision	78
Recall	60
f1_score	68

Table 5.13: Model performance excluding samples with mixed land cover: urban/vegetated.

6

ANALYSIS OF RESULTS

Random Forest (RF) achieved promising results with precision of 83%, recall of 83%, and f1-score of 83%, using bootstrap aggregation and optimized hyperparameters such as 50 decision trees, class weights of 1:5 (higher weight to the minority class), random selection of features at each split, and a maximum tree depth of 40. One of the main strategies to train the model and obtain acceptable performance results was to balance the dataset and work with a class ratio not greater than 1:14 (*landslide: non-landslides*). For this particular task, the merging algorithm (Section 5.2.3) reduced *non-landslide* segments from 41,529 to $\approx 2,749$ with a merging error of 8%. In most cases, we found that the merging error is caused by mislabeling during dataset preparation (Section 5.1.1) due to changes in image illumination, shadows, high sedimentation rate, cloud contamination and presence of small-scale landslides that were difficult to identify manually.

6.1 SEGMENTATION STRATEGIES

Dataset pre-processing allowed to generate cloud-free pre- and post-landslide images and compute band ratios and image differences (e.g. RGD, NDVI, VID) (Figure 6.1) that were key inputs for the two-step segmentation strategy.

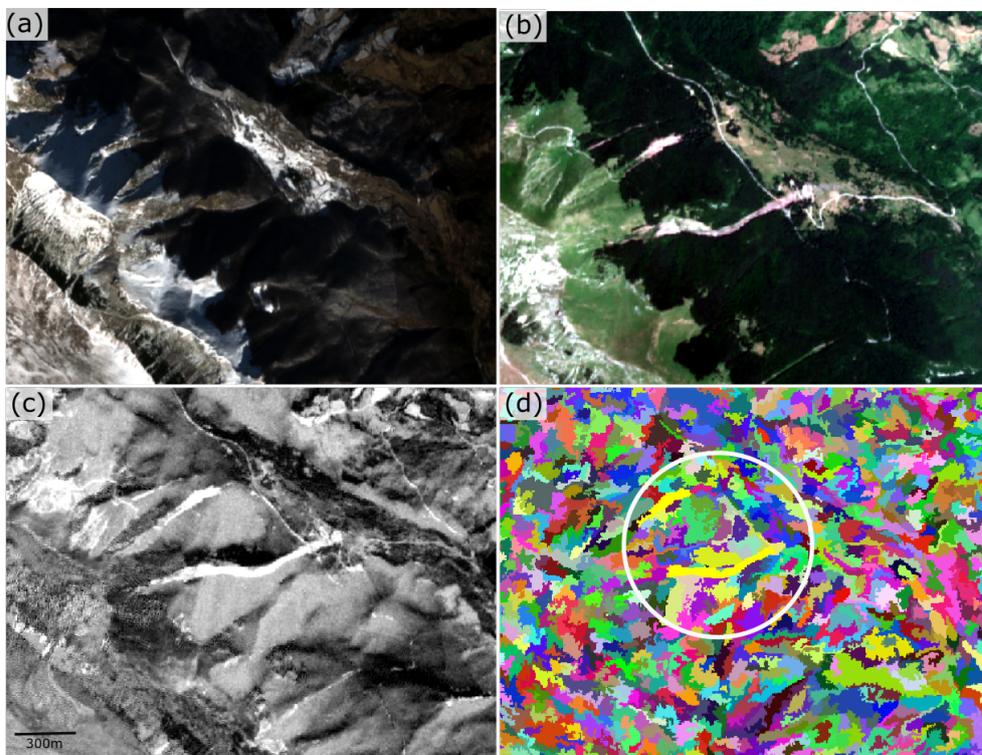


Figure 6.1: Image pre-processing and segmentation; sample in a remote area in Italy (L17). (a) Cloud-free pre-landslide image. (b) Cloud-free post-landslide image. (c) Image difference using band ratioing red/green (RGD). (d) Image segmentation.

K-means using the Red/Green Difference (RGD) as input feature and a k value of 19 was the most appropriate approach to identify landslides of different sizes and with variable spectral information. The applied segmentation strategy allowed to detect and isolate changes in pixels reflectances that can be likely related to the occurrence of landslides events. As an example, Figure 6.1 shows large-scale landslides occurring in a remote area in Italy. Although the larger landslide is contaminated with vegetation, the applied segmentation correctly delineated the borders of the landslide. The figure also shows how pixel features grouped and average into segments allowed to reduce the effect of isolated pixels and the problem of the class mixture within the same object.

The initial segmentation with fixed parameters across the entire dataset leads to oversegmentation of *non-landslide* areas, and consequently to an imbalanced dataset with a class ratio of 1:225 favoring the negative class. As RF and most of ML classifiers are biased toward the majority class, an initial assessment of the model performance using the above mentioned class distribution only reached a precision, recall, and f1-score of 71%, 55%, and 62%.

6.1.1 Optimization of the number of segments

An optimization strategy that reduces the oversegmentation of *non-landslide* areas was applied to balance the sample dataset and improve model performance. It consists in a merging algorithm that groups homogeneous segments based on a similarity criterion that is adapted to the distribution of the spectral features in each satellite image (Figure 5.7). This strategy reduces the risk of merging segments from different classes when compared to the use of a fixed threshold for the entire dataset. The challenge was to find a trade-off or balance point between the segments reduction and the merging error that can lead to an optimal class distribution with minimum negative impact on the RF performance.

Applying the merging algorithm with a sequential merging strategy that uses NDVI, VID, and RGD distributions, and a fine-grained NDVI threshold, produced the best results with a merging error of 8%, and *non-landslide* segments reduction of 93.3% (Table 5.9). The fine-grained threshold of 0.05 helped to merge homogeneous areas around the landslides avoiding the mixture of classes.

Other strategies as using a fine-grained fixed NDVI threshold (e.g. 0.08) without considering the distribution of spectral features, did not reduce the segments to an optimal number that allowed to achieve promising results using RF classification.

An important consideration when applying the aforementioned merging strategies is the fact that even if the goal is to maximize the merging of *non-landslide* segments, a minimum degree of heterogeneity should be considered in order to avoid the generation of segments that could resemble the landslides properties. Examples are segments belonging to urban areas (e.g. NDVI = 0.2) that could be merged with segments representing grasslands or croplands (e.g. NDVI = 0.6). Assuming that both classes cover the same area, the new NDVI values after merging the segments will be around 0.4, a value that could approximate the NDVI of landslides contaminated with vegetation.

6.2 MODEL TRAINING AND CLASSIFICATION

A total of 2,749 *non-landslide* segments are generated from the optimization strategy applied in previous section (Section 6.1.1). After cleaning outliers (Section 5.3.1), this number was reduced to 2,706 segments resulting in a class ratio of 1:14; the minimum optimal class distribution needed to retain a precision and recall over 60%. With this new class distribution (*landslides*=199, *non-landslides*=2,706), the RF classifier was trained using 70% of the sample dataset that corresponds to 1,894

non-landslides and 139 *landslides*. The iterative inclusion of selected *landslides diagnostic features*, progressively increased the model performance from precision, recall, and f1-score of 18%, 17%, and 17% to 85%, 61%, and 71% (Figure 5.10). The first criterion to discriminate *landslide* segments candidates from *non-landslide* was the *NDVI* (Table 5.10). It has been widely used as the most important *landslides diagnostic features* in previous works; however, as illustrated in Figure 5.10, this feature itself did not give enough information to detect landslides; thus, the model needs addition features to reach a good performance. Looking at Figure 6.2, it can be suggested that when working with different type of landslides, and occurring in a large variety of land cover, landscapes and *geomorphological settings*, *NDVI* can have a wide range of values (from 0 to 0.5) due to factors such as contamination by vegetation, and variation in soil composition and soil moisture. Hence, while more variation in the sample set, the higher is the probability to find *non-landslide* segments with similar *NDVI* values than *landslides*.

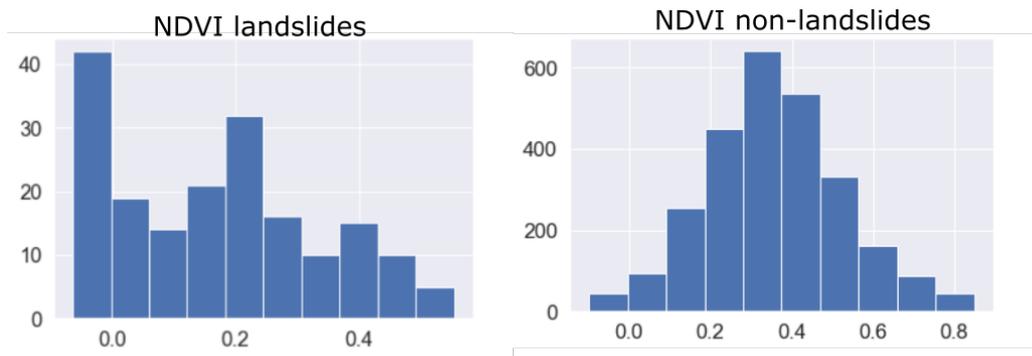


Figure 6.2: NDVI histograms for *landslide* and *non-landslide* segments.

Using bootstrap sampling, an experimental value of 50 decision trees, and other default tuning hyperparameters, features were iteratively added following the order shown in Table 5.10. It can be noticed that the inclusion of features related to changes such as $RGD_{deviation}$ and $BrightnessD_{deviation}$, importantly increased the model performance from precision, recall, and f1-score of 18%, 17%, and 17%, to 61%, 37%, and 46% (Figure 5.10). These two properties were key to discriminate landslides from objects with similar *NDVI* values that were not induced by changes during the same time of the landslide occurrence, including deforested areas, fresh rock, or older landslides events. Another measure of change, the $VID_{deviation}$ is the fourth feature input to the model; it achieved a considerable increase in the results to 79%, 56%, and 65%. $VID_{deviation}$ helped to discriminate between changes induced by removal of vegetation (likely landslides) and those caused by vegetation growing. As Figure 5.10 shows, the remaining features tend to slightly increase the model performance; however, the higher achievement was made by the first four features added to the model.

The experiments also revealed that precision always remains higher than recall (Figure 5.10). It can be suggested that this trend is likely due to the limited number of *landslide* segments (139) used to train the model, and the imbalanced distribution (1:14) that still remains even after the segment optimization. Hyperparameters tuning assigning weights to classes, was an optimal strategy to increase the balancing between precision and recall measures.

6.2.1 Final model assessment

With all *landslides diagnostic features* depicted in Table 5.10 and after hyperparameters tuning, the model achieved a precision of 83%, recall of 83%, and f1-score of 83%, adding more weight (1:5) to the *landslide* class; i.e. assigning a higher penalty when misclassifying the minority (Section 2.5.2).

Looking at the confusion matrix depicted in Figure 5.11, it can be seen that from a total number of 65 landslides segments used for the testing, 54 were correctly identified (true positives), 11 were not identified (false negatives), and 11 were misidentified (false positives). It is important to note that, depending on the application, the cost of false positives (low precision) will be higher than the cost of false negatives (low recall). Varying some hyperparameters as the class weight allows to slightly increase or decrease such metrics. If a higher weight is assigned to the *landslide* class, the recall can be improved but at the cost of reducing precision.

The final ranking of the features importance using Mean Decrease Impurity (MDI) is illustrated in Figure 5.12. It indicates that features related to spectral changes were the most relevant to build the classification model. One breakthrough in the context of the landslide detection from optical satellite imagery was the used of *change features* (e.g. *RGD*, *BrightnessD* and *VID*) relative to the *contextual* information of the image ($RGD_{deviation}$, $BrightnessD_{deviation}$, $VID_{deviation}$) (Equation 3.13). In the presence of a dataset characterized by a great variability on the distributions of the spectral information, those features were crucial to achieving promising results.

Although we initially ranked the *NDVI* in the first position (Table 5.10), results from *MDI* indicate that it is the third more important feature for our model. Regarding the topographic features, the *Relative_relief* shows higher contribution than the *Slope_mean* and the *Slope_max*. Even though topographic features are less relevant than the spectral ones, their exclusion reduces the model performance in about 10%.

The experiment conducted in Section 5.3.2 to evaluate the impact of training samples with mix urban/vegetated areas demonstrated that these samples have a relevant contribution to the model as their exclusion decrease the performance. Figure 5.13 indicates that without those samples, the model increases considerably the number of false negatives resulting in a recall of 60%. The precision and *f1-score* were also negatively influenced, resulting in a precision of 70% and recall of 68%. From this experiment we can conclude that the use of training samples located in urban areas does not negatively impact the model as far as half of the landslide is bordered by vegetation, it has a minimum size of 10,000m (100m × 100m), and the image has a minimum quality of class *M3* (Table 5.3).

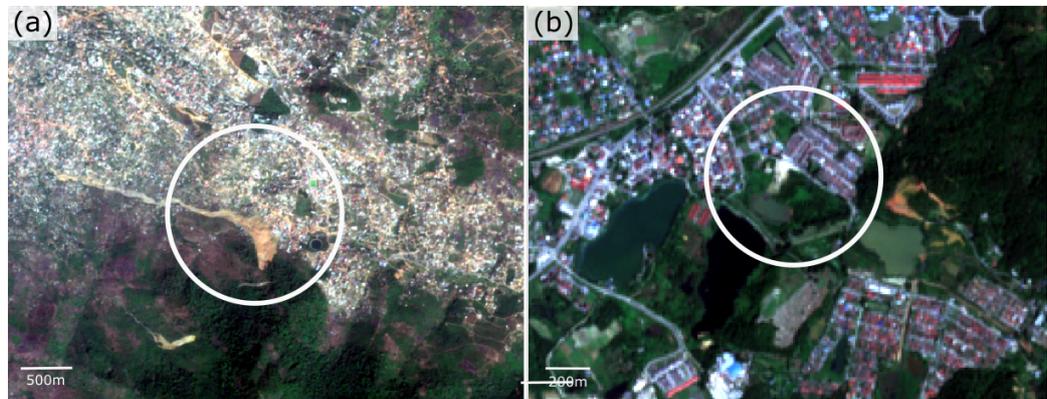


Figure 6.3: Training samples with mixed land cover: urban-vegetated.(a) Landslide *L2*. (b) Landslide *L7*.

6.3 VALIDATION

To validate the model we conducted the following experiments:

1. Validation on a satellite image with fresh landslides occurrences. Figure 6.4 illustrates the post-event satellite image (a) and the image difference (b). Regions circled in yellow were correctly identified as landslides, while those

areas circled in red were not detected. Analyzing the results, we found that the initial segmentation and merging algorithm separated *landslide* and *non-landslide* segments correctly. By looking at the segment features, we concluded that the non-detection of the landslide is due to the negative *NDVI* (-0.17) that the segment exhibits. These negative values can be related to the proximity of the segment to a creek, and consequently, to the presence of high water content. According to [Section 5.3.1](#), segments with $NDVI < -0.1$ are considered as outliers and therefore eliminated after the segmentation stage.

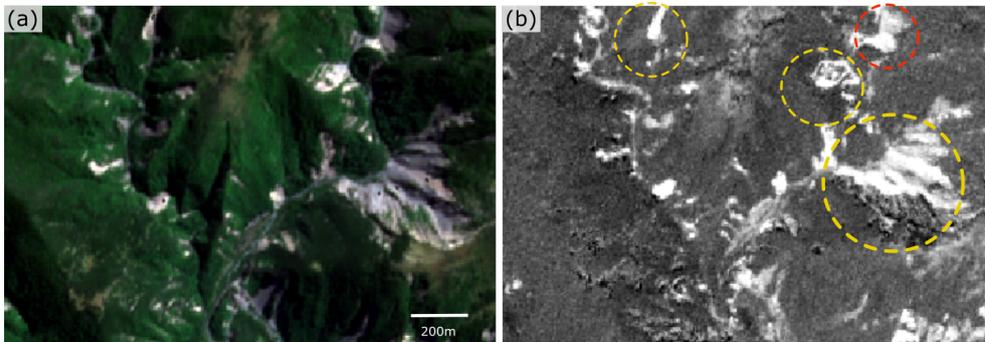


Figure 6.4: Model validation of landslides in New Zealand (event date: 2016-11-15). (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. Yellow circle, landslides detected. Red circle, landslides not detected.

- Validation on a satellite image that contains landslide that occurred before the acquisition time of the pre-image that we used to generate the image difference ([Equation 3.5](#)) ([Figure 6.5](#)). As expected, the model did not identify changes in [spectral features](#) values (low *VID*, *RGD*, *BrightnessD*); therefore, the landslide was ignored during the segmentation stage, and consequently during classification. This result is inline with the correct functionality of our classifier which is to label as *non-landslide*, those segments with low spectral changes.

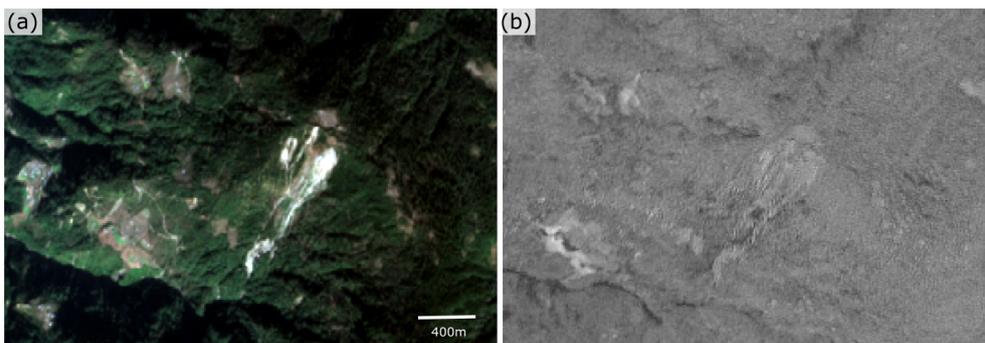


Figure 6.5: Model validation on area with old landslides. (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. No landslides detected.

- Validation on a satellite image from an area with non-probability of being affected by landslides occurrences ([Figure 6.6](#)). Even though several segments show similar [spectral features](#) values than landslides (high *RGD*, > 0.7 ; high *VID*, > 0.7 ; low *NDVI*, < 0.2), no landslides were detected. For this case, topographic features as slope and relative relief can be playing an important role in the model performance.

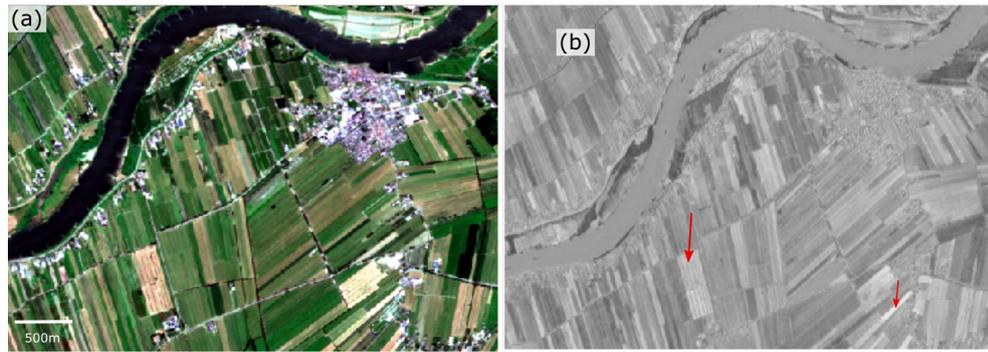


Figure 6.6: Model validation in the Netherlands. (a) Post-event image. (b) Image difference using RGD. White areas, high change between pre- and post image. Dark grey areas, low change. Red arrows indicate segments with similar spectral information than landslides. No landslides detected

4. Validation on a case study in Nepal where the NASA tool ¹ [Fayne et al., 2019] detected landslides (Figure 6.7). (a) Landslides detection using the NASA tool. (b) Image difference using RGD. (c) Post-event image. Our tool was able to detect one area with landslides that the NASA tool also identified (yellow circle). The area circled in red (detected by NASA as positive landslides) was not identified as a region affected by landslides. For this case, we found that our segmentation and merging algorithm separate *landslide* from *non-landslide* segments correctly; however, some *spectral features* of those segments show values that highly differ from the range of values in landslides. *BrightnessD* values (low brightness changes) close to 0 suggest that the landslides were likely reactivated or occurred before the acquisition time of the pre-image. The high value in *NDVI* suggests that those segments (circled in red) were contaminated with vegetation.

The NASA tool and our method for landslide detection use completely different approaches and have different scopes. In addition to spectral and topographic features, the NASA tool uses changes in soil moisture; therefore, the tool only works for rainfall-induced landslides and not for events caused by other triggers as earthquakes. In contrast to our method that is based on *ML* for image classification, the NASA tool uses *rule-based* approaches and thresholds that are region-specific.

¹ Sudden Landslide Identification Product (SLIP) from NASA is a software developed in Google Earth Engine that applies change-detection algorithms to identify landslides events. The tool uses features derived from optical satellite images, topographic slope, and soil moisture. It is based on rule-based approaches using region-specific thresholds that are calibrated from spectral analysis of several landslide events in Nepal. The model reported an overall accuracy of 56% over the entire study region with errors of commission (false positive) commonly resulting from newly cleared agricultural areas [Fayne et al., 2019].

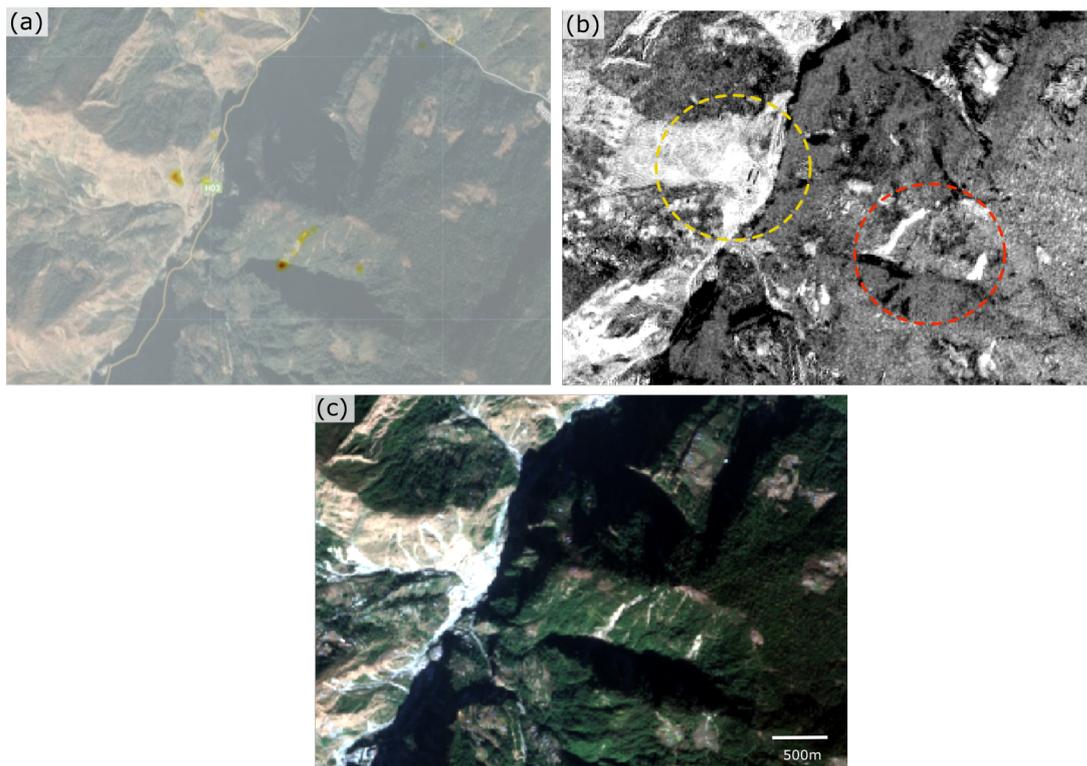


Figure 6.7: Model validation in Nepal. (a) Landslides detection using the NASA tool. (b) Image difference using RGD. (c) Post-event image. White areas, high change between pre- and post image. Dark grey areas, low change. Yellow circle, landslides detected for our tool. Red circle, landslides not detected.

7.1 CONCLUSIONS

The main research question of this study is *How to detect landslides using Machine Learning?* To answer this question we investigated a number of sub-questions:

1. *To what extent can landslides be detected using Sentinel-2 (GSD = 10m) in combination with a worldwide DEM (GSD = 30m)?*

The results demonstrated that landslides can be detected using features derived from the spectral information of Sentinel-2 images and topographic features from DEM with a precision of 83%, recall of 83%, and f1-score of 83%. Topographic features have less influence than spectral ones; however, their exclusion decreases the model performance to precision, recall, and f1-score of 74%, 71%, and 72%.

2. *What are the relevant landslide diagnostic features?*

The relevant features to detect landslides are illustrated in [Figure 5.12](#). Contextual features such as [BrightnessD_{deviation}](#) and [RGD_{deviation}](#) are the more relevant ones with an importance of 26% and 19%. [NDVI](#) and [VID_{deviation}](#) are the third and fourth more important features ranked with 13% and 12%. Each of the remaining features has less than 10% of importance, being the [slope_max](#) the one with less contribution to the model. For this study, [contextual](#) features were key features to overcome the challenge of working with satellite images distributed worldwide and with a wide variety in the statistical distributions of [spectral features](#).

3. *What is the best segmentation strategy?*

The best segmentation strategy is a two-step approach. In the first step, we use k-means (k=19) and the Red/Green difference ([RGD](#)) as the input feature to segment the image in homogeneous segments. For this segmentation, we used a criterion in which landslides are not merged with other segments. [RGD](#) was selected as the segmentation feature as it correctly highlights the changes induced by landslides and reduces variations caused by other irrelevant factors (e.g. changes in illumination).

The initial segmentation leads to oversegmentation of *non-landslide* areas and, consequently, to an imbalanced dataset. To overcome the oversegmentation challenge, we applied a merging algorithm using [NDVI](#) as homogeneity criterion to reduce the number of *non-landslide* segments to a class ratio not greater than 1:14 (*landslide* : *non-landslides*), which was the minimum class distribution required to keep a precision, recall and f1-score over 60%. Assuming that the vegetation is sensitive to the occurrences of landslides events due to the vegetation removal, [NDVI](#) was selected as the homogeneity criterion as it properly allows to discriminate *landslides* from *non-landslides* ([Section 3.2.2](#) and [Section 5.2.3](#)). To get the optimal class distribution with minimum negative impact on the [RF](#) performance, we experimentally found a trade-off or balance point between the segment reduction and the percentage error caused by wrongly merged landslides. This trade-off resulted in 93.3% of segment reduction (2,749 segments) with a merging error of 8%.

4. *How to exploit features per pixels to produce features per segments?*

Each segment was attributed with the features used to train the model (Section 5.2.2). For this, features at pixel level were grouped by computing statistical measures such as mean and maximum values within the segments polygons. Mean values were calculated for all features derived from Sentinel-2 and DEM. For the slope, the maximum value per segment was also computed (Table 5.7).

5. *What is the most appropriate Machine Learning technique?*

The most appropriate ML technique is the Random Forest classifier (Section 3.3). It has proven its effectiveness in a broad range of applications of remote sensing, especially for object-based image classification (Section 2.2.2). As a non-parametric method, the key property of Random Forest is its capability to handle different statistical distributions of features, which was one of the main challenges of this study. Furthermore, its capability to handle imbalanced datasets and missing values, as well as the use of few tuning parameters and not complex implementation, make Random Forest the most suitable algorithm for our binary classification problem.

6. *What is the accuracy of the most appropriate ML technique?*

The built Random Forest model achieved a precision 83%, recall of 83% , and f1-score of 83% for the *landslide* class. The confusion matrix depicted in Figure 5.11, indicates that from 65 landslides, 54 were correctly identified as true positives, 11 incorrectly detected as false positive, and 11 were missed or not identified.

This study achieved promising results for assisting the detection of landslides and optimizing time- and cost- consuming mapping routines. It is the first attempt of a method trained with satellite images from different areas, landscapes, and geomorphological settings and with different types of landslides not tied to a specific area and triggering factor.

The developed method is based on the principles OBIA and comprises a set of pre-processing and processing steps, including image acquisition or generation of cloud-free images from Sentinel-2, the image segmentation using spectral and topographic features extracted from Sentinel-2 and DEM, and image classification using Random Forest. Our method is built using entirely open source technologies allowing its applicability, re-usability, testing, and improvement (https://github.com/mhscience/landslides_detection).

Our method demonstrated not to be affected by areas with mixed urban-vegetation land cover as far as half of the landslide is bordered by vegetation, it has a minimum size of 10,000 m² and the image has a higher quality than M₃ (Section 5.1.1, Table 5.3). The method has limited performance in areas with perennial snow, high sedimentation rates, regions characterized by sparse or non-vegetation.

7.2 FUTURE WORKS

As a continuation of this study we recommend the following:

- Use our method to detect new landslides distributed worldwide and include them as part of the training samples. A higher number of landslides will increase the robustness and generalization of the model. Additionally, it will allow the comparison with other ML algorithms as Convolutional Neural Networks.
- In addition to the merging algorithm, we propose to explore a complementary approach to reduce the number of *non-landslide* segments. One strategy can be

to evaluate the segment redundancy and eliminate those segments that share similar spectral information (e.g. similar [NDVI](#), [GNDVI](#), [RGD](#)). This strategy will help to balance the dataset and keep accurate classification results for the *landslide* class while new training samples are added to the model.

- Research the inclusion of Synthetic Aperture Radar (SAR) images in our processing method to add terrain deformation features from pre-event and post-events. SAR signal penetration could overcome the limitation of optical images in regions with perennial snow, sand cover, and sparse or non-vegetation.
- Explore a strategy to include an automatic and dynamic [NDVI](#) threshold that instead of being experimentally derived, can be automatically calculated based on the spectral information contained in every image.

BIBLIOGRAPHY

- Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation*, 9(7):1545–1588.
- Baatz, M. (2000). Multi resolution segmentation: an optimum approach for high quality multi scale image segmentation. In *Beutrage zum AGIT-Symposium. Salzburg, Heidelberg, 2000*, pages 12–23.
- Bannari, A., Morin, D., Bonn, F., and Huete, A. (1995). A review of vegetation indices. *Remote sensing reviews*, 13(1-2):95–120.
- Baraldi, A., Puzzolo, V., Blonda, P., Bruzzone, L., and Tarantino, C. (2006). Automatic spectral rule-based preliminary mapping of calibrated landsat tm and etm+ images. *IEEE Transactions on geoscience and remote sensing*, 44(9):2563–2586.
- Behling, R., Roessner, S., Kaufmann, H., and Kleinschmit, B. (2014). Automated spatiotemporal landslide mapping over large areas using rapideye time series data. *Remote Sensing*, 6(9):8026–8055.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58(3-4):239–258.
- Berlind, R. S. (1995). An alternative method of stochastic discrimination with applications to pattern recognition.
- Bholowalia, P. and Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).
- Biau, G. and Scornet, E. (2016). A random forest guided tour. *Test*, 25(2):197–227.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16.
- Blaschke, T. (2013). Object based image analysis: A new paradigm in remote sensing. In *ASPRS Annual Conference, March*, pages 24–28.
- Blaschke, T., Burnett, C., and Pekkarinen, A. (2004). Image segmentation methods for object-based analysis and classification. In *Remote sensing image analysis: Including the spatial domain*, pages 211–236. Springer.
- Blaschke, T., Feizizadeh, B., and Hölbling, D. (2014). Object-based image analysis and digital terrain analysis for locating landslides in the urmia lake basin, iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4806–4817.
- Blaschke, T., Lang, S., Lorup, E., Strobl, J., and Zeil, P. (2000). Object-oriented image processing in an integrated gis/remote sensing environment and perspectives for environmental applications. *Environmental information for planning, politics and the public*, 2:555–570.
- Bostrom, H. (2007). Estimating class probabilities in random forests. In *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, pages 211–216. IEEE.

- Branco, P., Torgo, L., and Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR)*, 49(2):31.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.
- Breiman, L. (2001a). Random forests. *Machine learning*, 45(1):5–32.
- Breiman, L. (2001b). Random forests. *Machine learning*, 45(1):5–32.
- Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*, 1.
- Breiman, L., Freidman, H., Olshen, A., and Stone, C. (1984). Classification and regression trees. monterey ca: Wadsworth & brooks/cole advanced books & software.
- Briem, G. J., Benediktsson, J. A., and Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remote sensing data. *IEEE transactions on geoscience and remote sensing*, 40(10):2291–2299.
- Bunting, P., Clewley, D., Lucas, R. M., and Gillingham, S. (2014). The remote sensing and gis software library (rsgislib). *Computers & geosciences*, 62:216–226.
- Camps-Valls, G. (2009). Machine learning in remote sensing data processing. In *2009 IEEE international workshop on machine learning for signal processing*, pages 1–6. IEEE.
- Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Rojo-Álvarez, J. L., and Martínez-Ramón, M. (2008). Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1822–1835.
- Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., and Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. *IEEE geoscience and remote sensing letters*, 3(1):93–97.
- Carleer, A., Debeir, O., and Wolff, E. (2005). Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering & Remote Sensing*, 71(11):1285–1294.
- Castilla, G. and Hay, G. (2008). Image objects and geographic objects. In *Object-based image analysis*, pages 91–110. Springer.
- Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y. (2015). Pcanet: A simple deep learning baseline for image classification? *IEEE transactions on image processing*, 24(12):5017–5032.
- Chen, C., Liaw, A., Breiman, L., et al. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, 110:1–12.
- Chen, C. H. and Ho, P.-G. P. (2008). Statistical pattern recognition in remote sensing. *Pattern Recognition*, 41(9):2731–2741.
- Chen, G., Hay, G. J., Carvalho, L. M., and Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, 33(14):4434–4457.
- Cheng, K., Wei, C., and Chang, S. (2004). Locating landslides using multi-temporal satellite images. *Advances in Space Research*, 33(3):296–301.
- Clewley, D., Bunting, P., Shepherd, J., Gillingham, S., Flood, N., Dymond, J., Lucas, R., Armston, J., and Moghaddam, M. (2014). A python-based open source system for geographic object-based image analysis (geobia) utilizing raster attribute tables. *Remote Sensing*, 6(7):6111–6135.

- Collins, J. B. and Woodcock, C. E. (1994). Change detection using the gramm-schmidt transformation applied to mapping forest mortality. *Remote Sensing of Environment*, 50(3):267–279.
- Coppin, P. R. and Bauer, M. E. (1996). Digital change detection in forest ecosystems with remote sensing imagery. *Remote sensing reviews*, 13(3-4):207–234.
- Cover, T. M., Hart, P., et al. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.
- Dalla Mura, M., Benediktsson, J. A., Waske, B., and Bruzzone, L. (2010). Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747–3762.
- Danneels, G., Pirard, E., and Havenith, H.-B. (2007). Automatic landslide detection from remote sensing images using supervised classification methods. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 3014–3017. IEEE.
- Debeir, O. (2001). *Segmentation Supervisée d'Images*. PhD thesis, Ph. D., Faculté des Sciences Appliquées, Université Libre de Bruxelles.
- Del Frate, F., Pacifici, F., Schiavon, G., and Solimini, C. (2007). Use of neural networks for automatic classification from high-resolution images. *IEEE transactions on geoscience and remote sensing*, 45(4):800–809.
- Deng, H. and Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition*, 46(12):3483–3489.
- Dhanachandra, N., Manglem, K., and Chanu, Y. J. (2015). Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764–771.
- Dougherty, G. (2012). *Pattern recognition and classification: an introduction*. Springer Science & Business Media.
- Duro, D. C., Franklin, S. E., and Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 hrg imagery. *Remote Sensing of Environment*, 118:259–272.
- Enderle, D. I. and Weih Jr, R. C. (2005). Integrating supervised and unsupervised classification methods to develop a more accurate land cover classification. *Journal of the Arkansas Academy of Science*, 59(1):65–73.
- ESA (2013). ESAsentinel online.
- Espindola, G., Câmara, G., Reis, I., Bins, L., and Monteiro, A. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, 27(14):3035–3040.
- Fawagreh, K., Gaber, M. M., and Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal*, 2(1):602–609.
- Fayne, J. V., Ahamed, A., Roberts-Pierel, J., Rumsey, A. C., and Kirschbaum, D. (2019). Automated satellite-based landslide identification product for nepal. *Earth Interactions*, 23(3):1–21.
- Feizizadeh, B., Blaschke, T., Tiede, D., and Moghaddam, M. H. R. (2017). Evaluating fuzzy operators of an object-based image analysis for detecting landslides and their changes. *Geomorphology*, 293:240–254.

- Froude, M. J. and Petley, D. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18:2161–2181.
- Fu, K. (1983). Pattern recognition techniques in remote sensing data analysis. *Proceedings of the Indian Academy of Sciences Section C: Engineering Sciences*, 6(2):153–175.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.
- Grabowski, S., Jóźwik, A., and Chen, C. (2003). Nearest neighbor decision rule for pixel classification in remote sensing. In *Frontiers of Remote Sensing Information Processing*, pages 315–327. World Scientific.
- Ham, J., Chen, Y., Crawford, M. M., and Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501.
- Haque, A. (2017). Band ratioing and differencing. University Lecture.
- Haralick, R. M. and Shapiro, L. G. (1985). Image segmentation techniques. In *Applications of Artificial Intelligence II*, volume 548, pages 2–10. International Society for Optics and Photonics.
- Hay, G. and Castilla, G. (2006). Object-based image analysis: strengths, weaknesses, opportunities and threats (swot). In *Proc. 1st Int. Conf. OBLA*, pages 4–5.
- He, H. and Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9):1263–1284.
- Herrera, M. (2018). A worldwide geo-database for landslide prediction. In *Internship Report, Deltares/Delft University of technology*.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE.
- Hölbling, D., Friedl, B., and Eisank, C. (2015). An object-based approach for semi-automated landslide change detection and attribution of changes to landslide classes in northern taiwan. *Earth Science Informatics*, 8(2):327–335.
- Hu, W., Huang, Y., Wei, L., Zhang, F., and Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015.
- Inglada, J. (2007). Automatic recognition of man-made objects in high resolution optical remote sensing images by svm classification of geometric image features. *ISPRS journal of photogrammetry and remote sensing*, 62(3):236–248.
- Ingram, K., Knapp, E., and Robinson, J. (1981). Change detection procedure development for improved urbanized area delineation. *Technical memorandum CSC/TM-81*, 6087.
- Iwahashi, J. and Pike, R. J. (2007). Automated classifications of topography from dems by an unsupervised nested-means algorithm and a three-part geometric signature. *Geomorphology*, 86(3-4):409–440.
- Jackson, Q. and Landgrebe, D. A. (2002). Adaptive bayesian contextual classification based on markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2454–2463.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

- Japkowicz, N. and Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449.
- JAXA (2017). Alos global digital surface model (dsm).
- Jensen, R. R., Hardin, P. J., and Yu, G. (2009). Artificial neural networks and remote sensing. *Geography Compass*, 3(2):630–646.
- Johnson, B. and Xie, Z. (2011). Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):473–483.
- Kansakar, P. and Hossain, F. (2016). A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth. *Space Policy*, 36:46–54.
- Kersbergen, D. (2018). Automated building damage classification using remotely sensed data: Case study: Hurricane damage on st. maarten.
- Keyport, R. N., Oommen, T., Martha, T. R., Sajinkumar, K., and Gierke, J. S. (2018). A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images. *International journal of applied earth observation and geoinformation*, 64:1–11.
- Khorram, S., Van Der Wiele, C. F., Koch, F. H., Nelson, S. A., and Potts, M. D. (2016). *Principles of applied remote sensing*. Springer.
- Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards*, 52(3):561–575.
- Kleinberg, E. (1990). Stochastic discrimination. *Annals of Mathematics and Artificial intelligence*, 1(1):207–239.
- Koprinska, I., Deng, D., and Feger, F. (2006). Image classification using labelled and unlabelled data. In *2006 14th European Signal Processing Conference*, pages 1–5. IEEE.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Lam, L. and Suen, C. Y. (1994). A theoretical analysis of the application of majority voting to pattern recognition. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)*, volume 2, pages 418–420. IEEE.
- Lang, S. (2008). Object-based image analysis for remote sensing applications: modeling reality—dealing with complexity. In *Object-based image analysis*, pages 3–27. Springer.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10.
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., and Bégué, A. (2017). A combined random forest and obia classification scheme for mapping small-holder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, vhrs and dem). *Remote Sensing*, 9(3):259.
- Lee, S. and Lee, M.-J. (2006). Detecting landslide location using kompsat 1 and its application to landslide-susceptibility mapping at the gangneung area, korea. *Advances in Space Research*, 38(10):2261–2271.

- Lemmens, M. (1994). Statistical pattern recognition of remotely sensed images. In *Delft University of technology, Faculty of Geodetic Engineering, Laboratory of Photogrammetry and Remote Sensing*.
- Lemmens, M. (2011). *Geo-information: technologies, applications and the environment*, volume 5. Springer Science & Business Media.
- Lemmens, M. (2015). Sentinel-2a in orbit. *GIM INTERNATIONAL-THE WORLD-WIDE MAGAZINE FOR GEOMATICS*, 29(7):22–25.
- Li, M., Ma, L., Blaschke, T., Cheng, L., and Tiede, D. (2016). A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *International Journal of Applied Earth Observation and Geoinformation*, 49:87–98.
- Li, M., Zang, S., Zhang, B., Li, S., and Wu, C. (2014). A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, 47(1):389–411.
- Liaw, A., Wiener, M., et al. (2002). Classification and regression by random forest. *R news*, 2(3):18–22.
- Lu, D., Mausel, P., Brondizio, E., and Moran, E. (2004). Change detection techniques. *International journal of remote sensing*, 25(12):2365–2401.
- Lu, P., Stumpf, A., Kerle, N., and Casagli, N. (2011). Object-oriented change detection for landslide rapid mapping. *IEEE Geoscience and remote sensing letters*, 8(4):701–705.
- Ma, L., Li, M., Ma, X., Cheng, L., Du, P., and Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277–293.
- Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. (2016). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657.
- Mahmoodzadeh, H. (2007). Digital change detection using remotely sensed data for monitoring green space destruction in tabriz. *International Journal of Environmental Research*, 1(1):35–41.
- Mannan, B. and Ray, A. (2003). Crisp and fuzzy competitive learning networks for supervised classification of multispectral IRS scenes. *International Journal of Remote Sensing*, 24(17):3491–3502.
- Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J., and Kumar, K. V. (2010). Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology*, 116(1–2):24–36.
- Martha, T. R., Kerle, N., van Westen, C. J., Jetten, V., and Kumar, K. V. (2011). Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12):4928–4943.
- Martha, T. R., Kerle, N., Van Westen, C. J., Jetten, V., and Kumar, K. V. (2012). Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. *ISPRS journal of photogrammetry and remote sensing*, 67:105–119.
- McDermid, G. J. and Franklin, S. E. (1994). Spectral, spatial, and geomorphometric variables for the remote sensing of slope processes. *Remote Sensing of Environment*, 49(1):57–71.

- Melgani, F. and Bruzzone, L. (2002). Support vector machines for classification of hyperspectral remote-sensing images. In *IEEE international geoscience and remote sensing symposium*, volume 1, pages 506–508. IEEE.
- Moosavi, V., Talebi, A., and Shirmohammadi, B. (2014). Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by taguchi method. *Geomorphology*, 204:646–656.
- Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.
- Nguyen, T., Han, J., and Park, D.-C. (2013). Satellite image classification using convolutional learning. In *AIP Conference Proceedings*, volume 1558, pages 2237–2240. AIP.
- Nithya, S. E. and Prasanna, P. R. (2010). An integrated approach with gis and remote sensing technique for landslide hazard zonation. *International Journal of Geomatics and Geosciences*, 1(1):66–75.
- OpenCv (2019). Support vector machines.
- O’Connell, J., Bradter, U., and Benton, T. G. (2015). Wide-area mapping of small-scale features in agricultural landscapes using airborne remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 109:165–177.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222.
- Pal, N. R. and Pal, S. K. (1993). A review on image segmentation techniques. *Pattern recognition*, 26(9):1277–1294.
- Parker, O. P. (2013). *Object-based segmentation and machine learning classification for landslide detection from multi-temporal WorldView-2 imagery*. PhD thesis, San Francisco State University.
- Peña, J. M., Gutiérrez, P. A., Hervás-Martínez, C., Six, J., Plant, R. E., and López-Granados, F. (2014). Object-based image classification of summer crops with machine learning methods. *Remote Sensing*, 6(6):5019–5041.
- Pike, R. J. (1988). The geometric signature: quantifying landslide-terrain types from digital elevation models. *Mathematical geology*, 20(5):491–511.
- Platt, R. V. and Rapoza, L. (2008). An evaluation of an object-oriented paradigm for land use/land cover classification. *The Professional Geographer*, 60(1):87–100.
- Pradhan, B. and Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software*, 25(6):747–759.
- Pretorius, A., Bierman, S., and Steel, S. J. (2016). A meta-analysis of research in random forests for classification. In *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, pages 1–6. IEEE.
- Puissant, A., Rougier, S., and Stumpf, A. (2014). Object-oriented mapping of urban trees using random forest classifiers. *International Journal of Applied Earth Observation and Geoinformation*, 26:235–245.
- Qian, Y., Zhou, W., Yan, J., Li, W., and Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1):153–168.

- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.
- Rajan, S., Ghosh, J., and Crawford, M. M. (2008). An active learning approach to hyperspectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(4):1231–1242.
- Raspini, F., Ciampalini, A., Bianchini, S., Bardi, F., Di Traglia, F., Basile, G., and Moretti, S. (2016). Updated landslide inventory of the area between the furiano and rosmarino creeks (sicily, italy). *Journal of Maps*, 12(5):1010–1019.
- Robnik-Šikonja, M. (2004). Improving random forests. In *European conference on machine learning*, pages 359–370. Springer.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39.
- Salembier Clairon, P. J., Oliveras Vergés, A., and Garrido Ostermann, L. (1998). Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing*, 7(4):555–570.
- San, B., Sumer, E., and Gurcay, B. (2004). Comparison of band ratioing and spectral indices methods for detecting alunite and kaolinite minerals using aster data in biga region, turkey. In *Proceedings ISPRS*.
- Sekertekin, A., Marangoz, A., and Akcin, H. (2017). Pixel-based classification analysis of land use land cover using sentinel-2 and landsat-8 data. *Forest*, 80(78.13):82–71.
- Seyedhosseini, M. and Tasdizen, T. (2015). Disjunctive normal random forests. *Pattern Recognition*, 48(3):976–983.
- Shepherd, J. D., Bunting, P., and Dymond, J. R. (2019). Operational large-scale segmentation of imagery based on iterative elimination. *Remote Sensing*, 11(6):658.
- Singh, A. (1989). Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, 10(6):989–1003.
- Singh, S. and Gupta, P. (2014). Comparative study id3, cart and c4. 5 decision tree algorithm: a survey. *International Journal of Advanced Information Science and Technology (IJAIST)*, 27(27):97–103.
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., and Macomber, S. A. (2001). Classification and change detection using landsat tm data: when and how to correct atmospheric effects? *Remote sensing of Environment*, 75(2):230–244.
- Strahler, A. H. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote sensing of Environment*, 10(2):135–163.
- Stumpf, A. and Kerle, N. (2011). Object-oriented mapping of landslides using random forests. *Remote sensing of environment*, 115(10):2564–2577.
- Subhashini, L. and Premaratne, H. (2013). Landslide prediction using artificial neural networks.
- SUHET (2013). Sentinel-2 user handbook.
- Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H. (2014). Precise global dem generation by alos prism. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(4):71.
- Tehrani, F., Santinelli, G., and Herrera, M. (2019). A framework for predicting rainfall-induced landslides using machine learning. In *The XVII European Conference on Soil Mechanics and Geotechnical Engineering*.

- Tripoliti, E. E., Fotiadis, D. I., and Manis, G. (2013). Modifications of the construction and voting mechanisms of the random forests algorithm. *Data & Knowledge Engineering*, 87:41–65.
- Tsangaratos, P. and Ilia, I. (2014). A supervised machine learning spatial tool for detecting terrain deformation induced by landslide phenomena.
- Tsybmal, A., Pechenizkiy, M., and Cunningham, P. (2006). Dynamic integration with random forests. In *European conference on machine learning*, pages 801–808. Springer.
- Tzotsos, A. and Argialas, D. (2008). Support vector machine classification for object-based image analysis. In *Object-Based Image Analysis*, pages 663–677. Springer.
- van Oosterom, P. (1999). Spatial access methods. *Geographical information systems*, 1:385–400.
- Van Westen, C. (2016). National scale landslide susceptibility assessment for grenada. *Caribbean Handbook on Risk Information Management, World Bank GF-DRR, ACP-EU Natural Disaster Risk Reduction*.
- Van Westen, C., Van Asch, T. W., and Soeters, R. (2006). Landslide hazard and risk zonation—why is it still so difficult? *Bulletin of Engineering geology and the Environment*, 65(2):167–184.
- Veena, V., Sai, S. G., Tapas, R. M., Deepak, M., and Rama, R. N. (2016). Automatic detection of landslides in object-based environment using open source tools.
- Wang, Z., Jensen, J. R., and Im, J. (2010). An automatic region-based image segmentation algorithm for remote sensing applications. *Environmental Modelling & Software*, 25(10):1149–1165.
- Weih, R. C. and Riggan, N. D. (2010). Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(4):C7.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37.
- Yang, W., Wang, M., and Shi, P. (2012). Using modis ndvi time series to identify geographic patterns of landslides in vegetated regions. *IEEE Geoscience and Remote Sensing Letters*, 10(4):707–710.
- Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., and Schirokauer, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering & Remote Sensing*, 72(7):799–811.
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., and Pourghasemi, H. R. (2018). Landslide susceptibility modeling applying machine learning methods: A case study from longju in the three gorges reservoir area, china. *Computers & geosciences*, 112:23–37.
- Zhou, W. and Troy, A. (2009). Development of an object-based framework for classifying and inventorying human-dominated forest ecosystems. *International Journal of Remote Sensing*, 30(23):6343–6360.

A | PRE-PROCESSING

id	event_date	longitude	latitude	location_accuracy	landslide_trigger	landslide_size	country	# events	quality	land cover class	~ area (m2)	Source
L0	2017-03-25	-76.66247976	1.169677465	exact	downpour	very large	Colombia	1	M1	Vegetated areas/ Urban	172500	NASA Catalog
L1	2017-06-27	103.6529857	32.06849057	exact	continuous_rain	catastrophic	China	1	H1	Vegetated	2000000	NASA Catalog
L2	2017-08-14	-13.22985664	8.436115153	exact	unknown	very large	Sierra Leone	1	M1	Vegetated areas/ Urban	96900	Web resources
L3	2017-01-10	-65.46772261	-23.9134425	approximated	downpour	very large	Argentina	1	S1	Bare-lands	-	NASA Catalog
L4	2016-05-18	80.4319754	7.16080832	exact	monsoon	very large	Sri Lanka	2	M1	Vegetated areas	300000	NASA Catalog
L5	2016-06-15	-117.2679	56.2246	approximated	downpour	large	Canada	1	S3	Croplands	11880	NASA Catalog
L7	2016-11-27	101.613538	3.366921	exact	rain	large	Malasia	1	M2	Urban	11703	NASA Catalog
L9	2017-06-12	-119.749369	37.67866603	exact	unknown	large	United States	1	M2	Vegetated areas/ Urban	11700	NASA Catalog
L10	2017-01-09	-124.304631	43.20369447	approximated	rain	large	United States	1	S3	Vegetated areas	-	NASA Catalog
L11	2017-04-20	-73.32606486	43.4509461	exact	unknown	large	United States	1	M1	Croplands	8000	NASA Catalog
L12	2017-02-09	-43.42657247	-20.24799736	exact	mining	very large	Brasil	1	H3	Wetlands	135270	NASA Catalog
L13	2017-02-24	18.07942725	44.14733296	approximated	rain	very large	Bosnia and Hei	1	S3	Croplands	-	NASA Catalog
L15	2017-05-20	-121.4323838	35.86562803	exact	rain	very large	United States	1	H2	Vegetated	232300	NASA Catalog
L16	2017-04-02	-118.1639859	47.88398686	exact	downpour	large	United States	1	H3	Vegetated areas	10560	NASA Catalog
L17	2017-01-18	13.77821597	42.42921449	exact	earthquake	large	Italy	2	H3	Vegetated area	72000	NASA Catalog
L18	2017-07-03	167.9689724	-44.6885808	exact	rain	large	New Zealand	2	H3	Vegetated area	52440	NASA Catalog
L19	2017-04-07	73.62572743	40.77327756	exact	snowfall_snowmelt	large	Kyrgyzstan	2	H3	Vegetated area	32215	Nasa Catalog
L20	2017-04-09	111.8266482	-7.232228759	exact	downpour	large	Indonesia	1	M1	Vegetated area/croplands	70000	Nasa Catalog
L21	2017-07-06	130.9355164	33.3891756	exact	tropical_cyclone	large	Japan	5	M3	Vegetated areas	36000	Nasa Catalog
L22	2017-07-06	130.9002283	33.38442516	bbox center	tropical_cyclone	various	Japan	5	M3	Vegetated areas	7667	Nasa Catalog
L23	2017-07-06	130.7620809	33.39481989	bbox center	tropical_cyclone	large	Japan	4	H3	Vegetated areas	15040	Time series GEE
L25	2017-07-06	130.8207411	33.40312701	largest landslide	tropical_cyclone	various	Japan	3	M2	Vegetated areas	33040	Time series GEE
L26	2017-05-24	130.8113097	33.40309442	exact	tropical_cyclone	various	Japan	1	M2	Vegetated areas	25875	Time series GEE
L27	2017-05-24	131.1126142	33.44220602	exact	tropical_cyclone	various	Japan	1	M2	Vegetated areas	11904	Time series GEE
L28	2017-05-24	131.0896135	33.4850438	exact	tropical_cyclone	various	Japan	1	M2	Vegetated areas	29036	Time series GEE
L40	2018-04-11	130.8797709	33.5292582	exact (major event)	unknown	various	Japan	6	H3	Vegetated areas	38478	Time series GEE
L41	2018-04-11	130.8966053	33.43075427	exact	unknown	large	Japan	1	H3	Vegetated areas	16185	Web Resources
L43	2018-09-06	141.92972	42.71625817	bbox center	earthquake	various	Japan	34	M1	Vegetated areas /croplands	19425	Web Resources
L47	2018-09-06	141.9907587	42.71624191	bbox center	earthquake	various	Japan	13	M1	Vegetated areas /croplands	2420	Web Resources
L48	2018-09-06	142.0517974	42.71619312	bbox center	earthquake	various	Japan	1	M1	Vegetated areas /croplands	8174	Web Resources
L55	2016-11-15	173.8168708	-42.20224732	bbox center	earthquake	large	New Zealand	2	H3	Vegetated areas	38324	Web Resources
L56	2016-11-15	173.8168708	-42.15723341	bbox center	earthquake	various	New Zealand	5	H3	Vegetated areas	57720	Web Resources
L57	2016-11-15	173.756415	-42.11220322	bbox center	earthquake	various	New Zealand	1	H3	Vegetated areas	13200	Web Resources
L58	2016-11-15	173.8774124	-42.20223134	bbox center	earthquake	various	New Zealand	7	H1	Vegetated areas	391840	Web Resources
L59	2016-11-15	173.8774124	-42.15721743	bbox center	earthquake	various	New Zealand	2	H2	Vegetated areas	158952	Web Resources
L75	2018-06-19	85.89029585	27.90834619	exact	unknown	large	Nepal	1	H2	Vegetated areas	17350	Web Resources

Figure A.1: Landslides database. An example of the landslides tables stored in PostgreSQL. The original data includes additional information such as the length and width of the major landslide recognized in the landslide category (e.g. landslide, mudslide), landslide setting (e.g. natural slope, above river), location name (e.g. Mocoa, Colombia; Java, Indonesia), water body proximity (e.g. close to river, lake or coast).

B | IMAGE SEGMENTATION

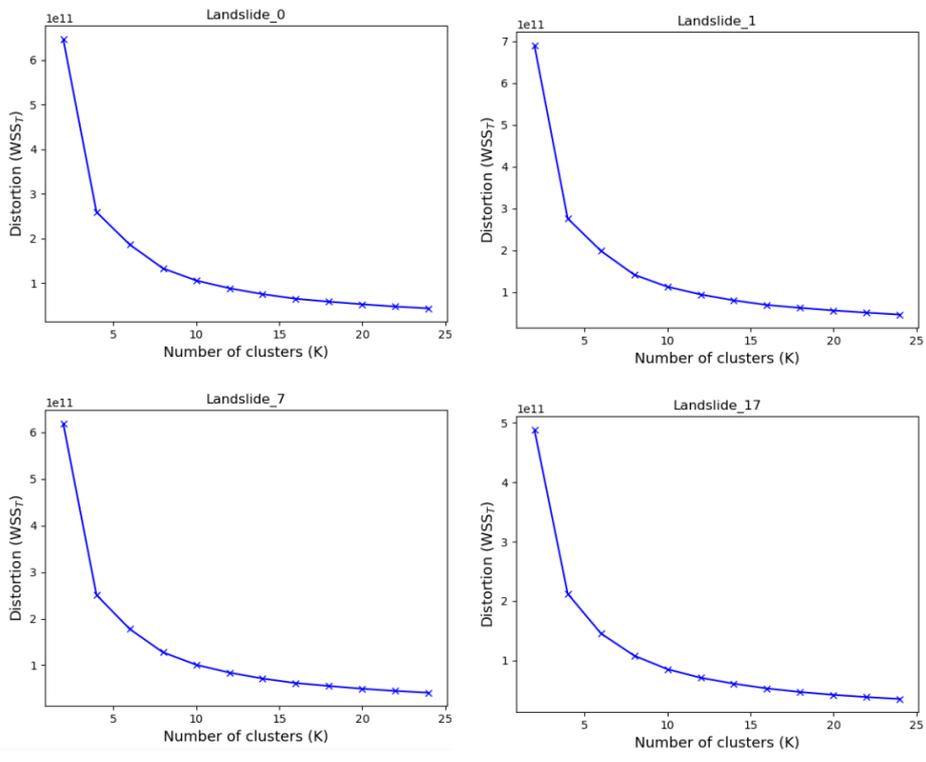


Figure B.1: Image set preparation.

C

EXPLORATORY ANALYSIS

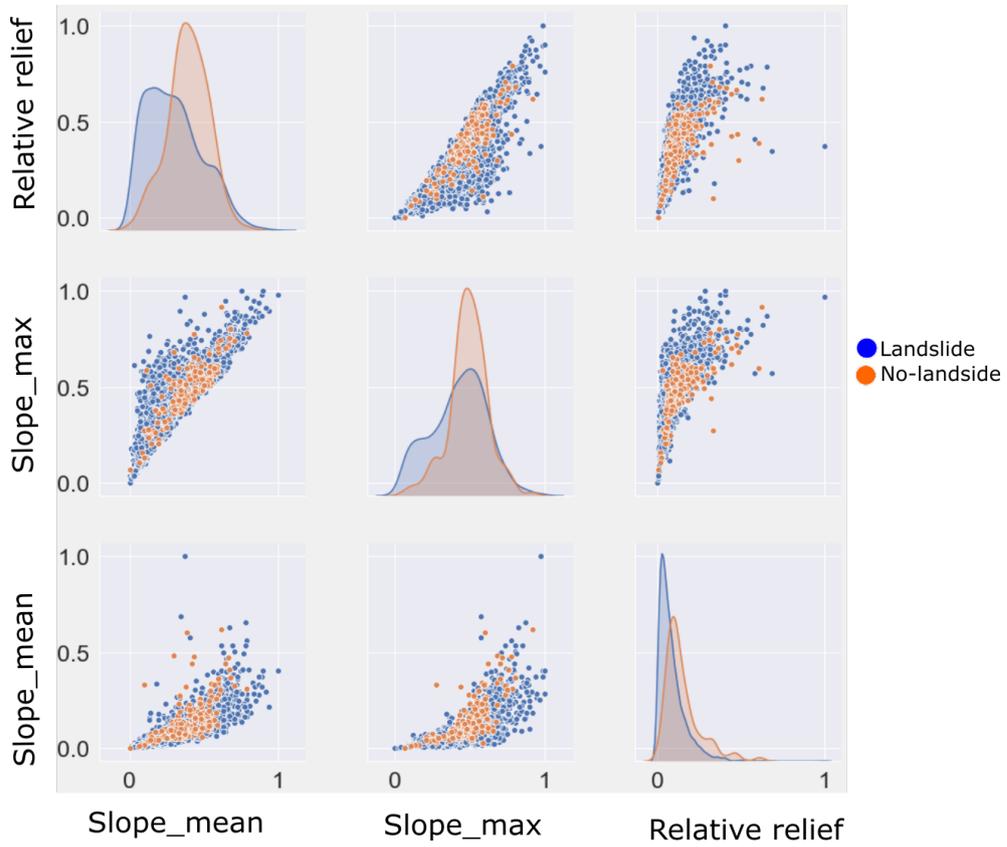


Figure C.1: Class distribution and correlation between features: Slope_mean, Slope_max, and Relative_relief.

COLOPHON

This document was typeset using \LaTeX . The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classithesis` package from André Miede.

