Landslide Detection using Random Forest classifier

Author: Meylin Herrera Herrera

P5 Final Thesis Presentation

Mentors

Dr.ir. Mathias Lemmens

Dr. ir. Amin Askarinejad

Ext. mentors

Dr. ir. Faraz Tehrani

Ir. Giorgio Santinelli

Outline

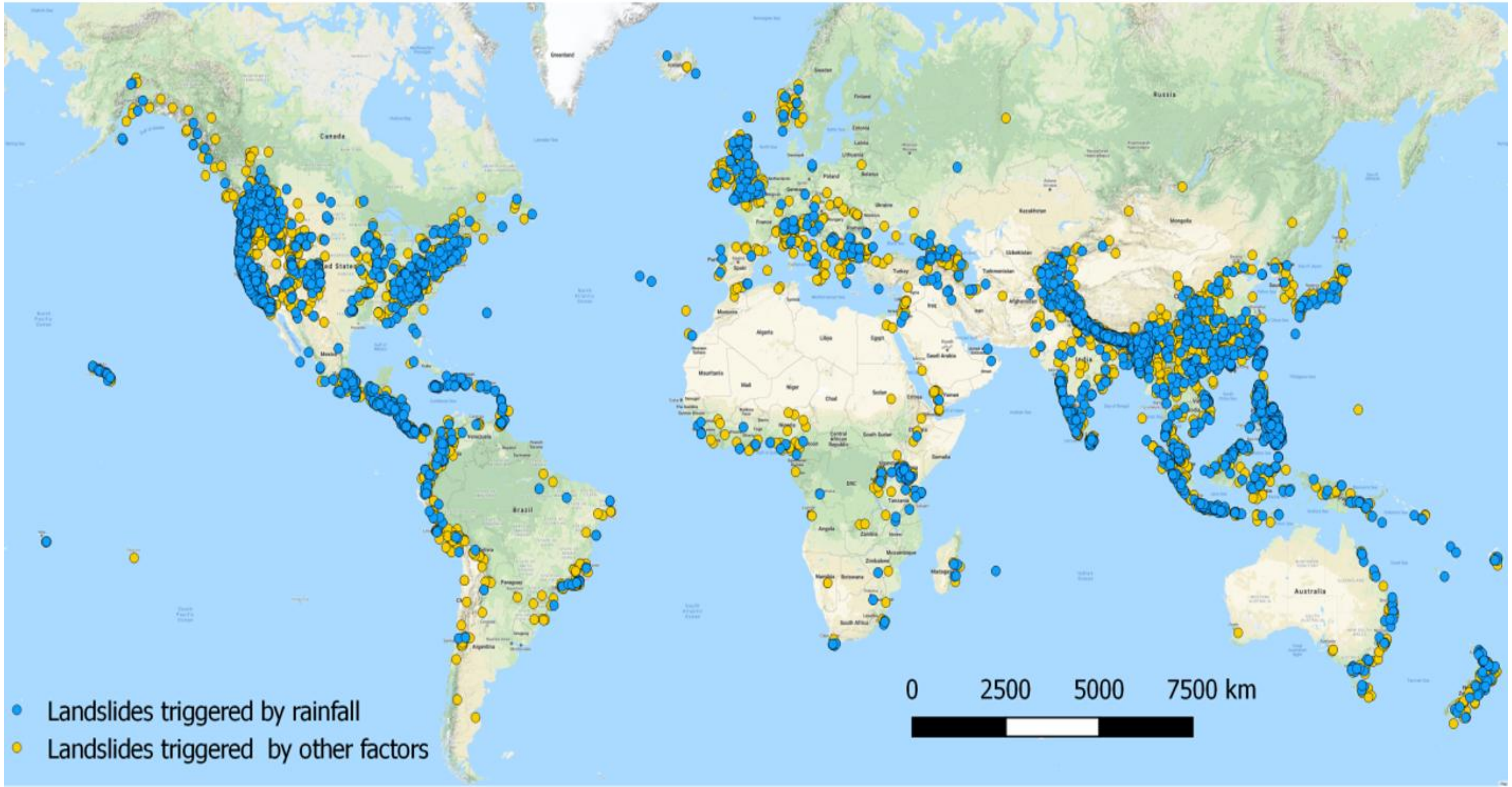
- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

Outline

- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

Motivation

NASA Global landslides Catalog (2007 – 2017)



~56.000 fatalities (2004-2016) (Froude and Petley, 2018)

Herrera, 2018

Motivation

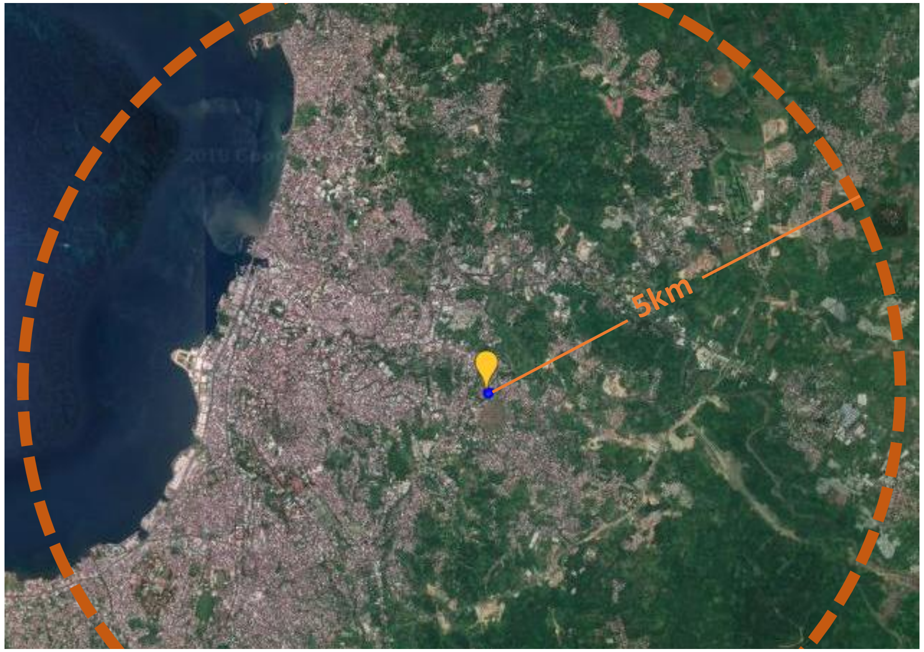

Japan, 2018[1]


New Zealand, 2016[2]

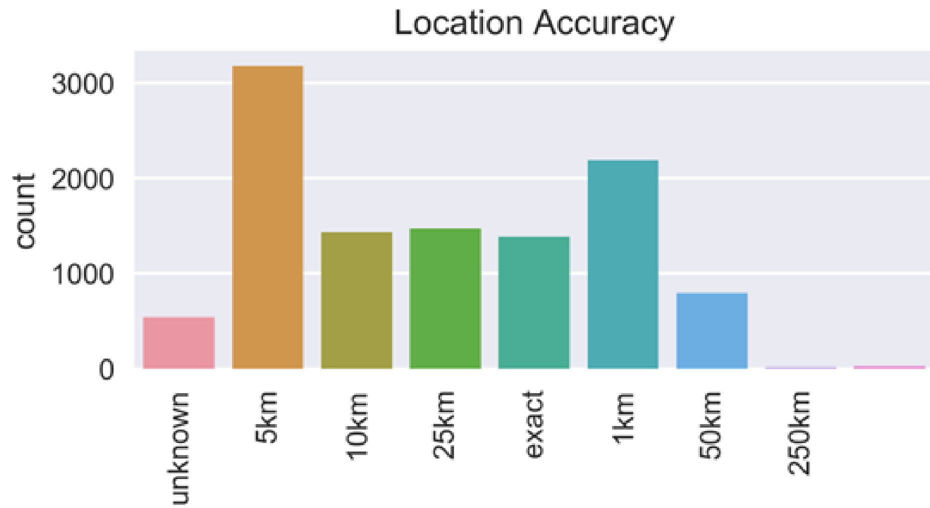

Sierra Leone, 2017[3]


Colombia, 2017[4]

Motivation



Case study Manado-Indonesia



Landslides location accuracy (Herrera, 2018)

Main problem — Landslide inventory maps (Accurate location)

Alternatives — Public Earth-observation data

How? — Machine Learning → automatic landslide detection

Research Questions

How to detect landslides using Machine Learning?

Research Questions

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

TU Delft

Deltares

New challenges

- Model generalization
- Multi-scale segmentation approach
- Applicability and re-usability

Outline

- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

TU Delft

Deltares

Image classification techniques

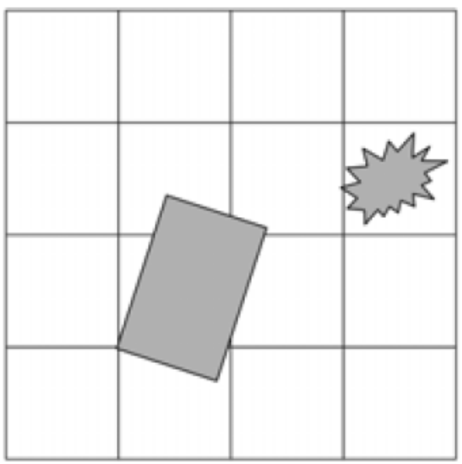
Pixel-based

- ✓ The **pixel** is the object
- ✓ Each **pixel** belongs to a class

Object-based (OBIA)

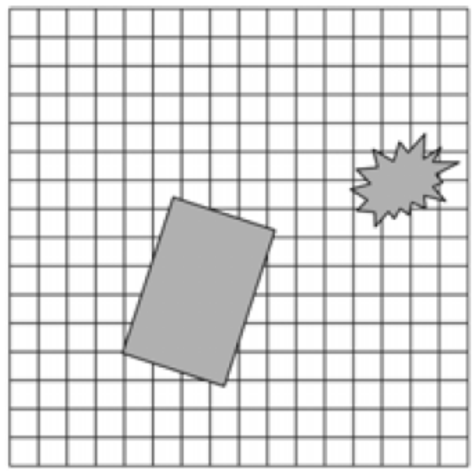
- ✓ The **segment** is the object
- ✓ Each **segment** belongs to a class

Medium Resolution



Blaschke (2010)

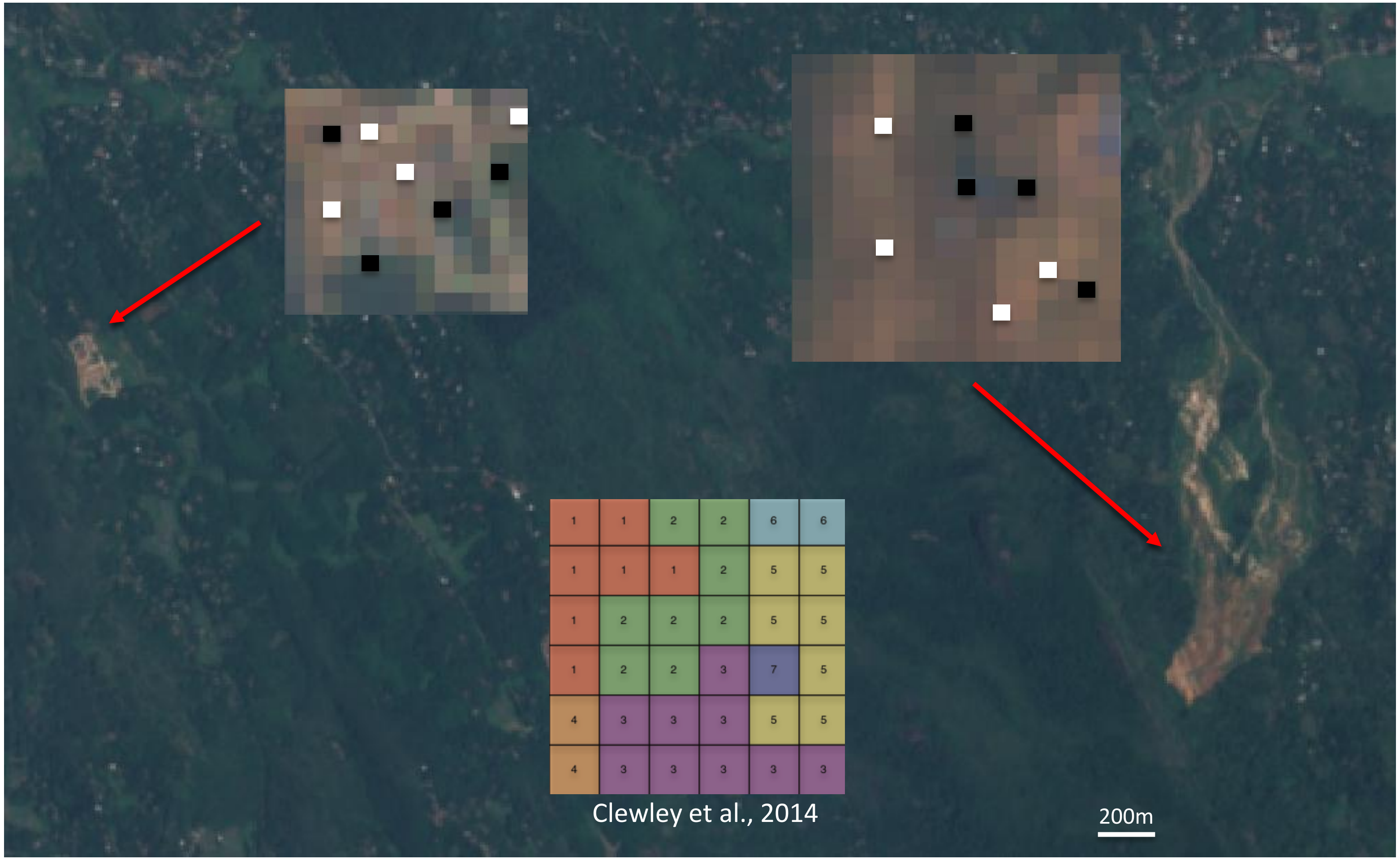
High to very high Resolution



Blaschke (2010)

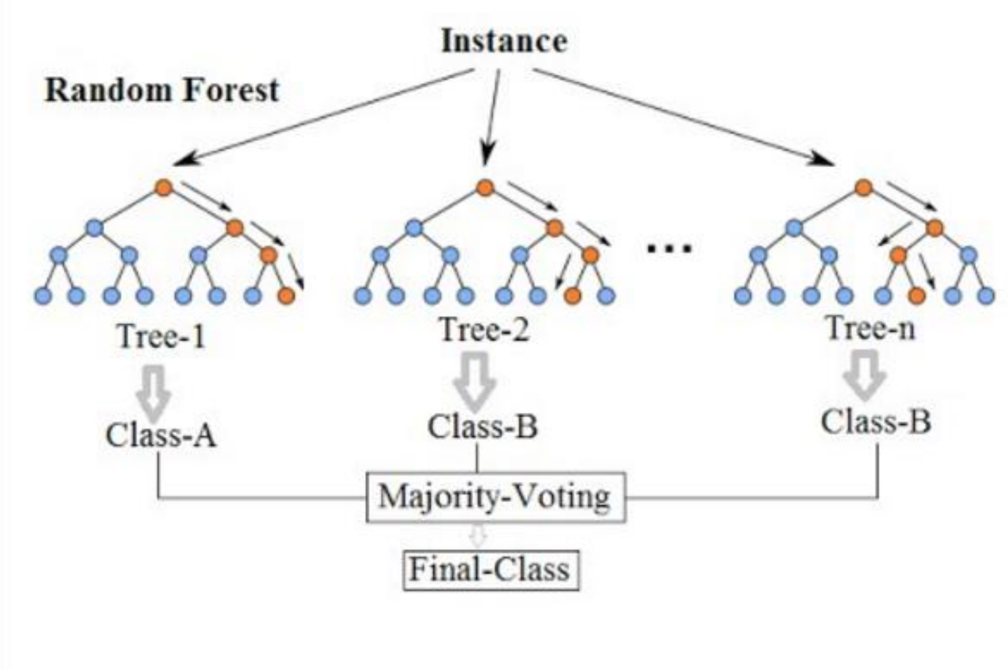
Why OBIA?

Spectral + spatial +contextual



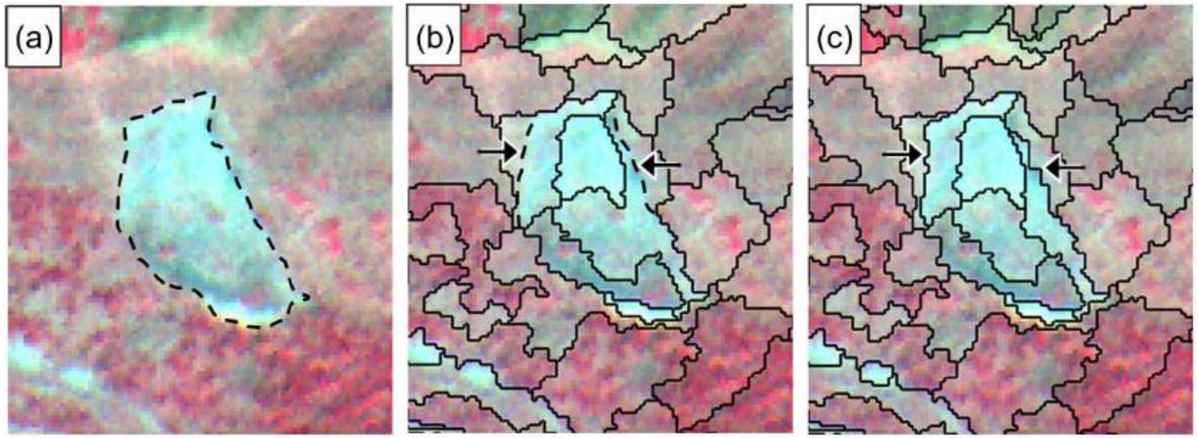
Supervised classification

Random Forest



Source image: https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page

OBIA in the context of landslides



Martha et al. (2011)

Object-based classification	Author
Rule-based	Martha et al. (2011)
Rule-based	Blaschke et al. (2014)
Rule-based	Holbling et al.(2015)
Machine Learning	Stumpf and Kerle (2011)
Machine Learning	Parker (2013)

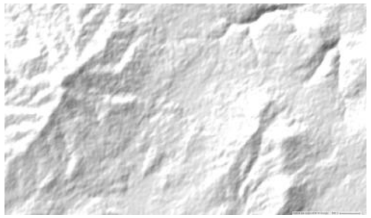
Outline

- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

Methodology overview

2015-present

Sentinel-2
(10m)



2006-2011

DEM ALOS
(30m)

Free-of-charge datasets

Pre-processing

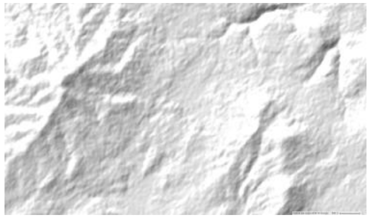
Image segmentation

Image classification

Methodology overview

2015-present

Sentinel-2 (10m)



2006-2011

DEM ALOS (30m)

Free-of-charge datasets

Pre-processing

(a)

Features computation at pixel level

Sample set preparation

Generation of cloud-free images

Extraction of pixel signatures

Band ratioing

Image normalization

Image difference

Image segmentation

Image classification

Sample set preparation

Landslides events: 110

Satellite Images: 96 (32 post-event / 32 pre-event / 32 image difference)

Sample set preparation

Quality assessment

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

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

Scale
H1
H2
H3
M1
M2
M3
S1
S2
S3

Image set preparation

Processed images (GEE) > 1,500
~1TB

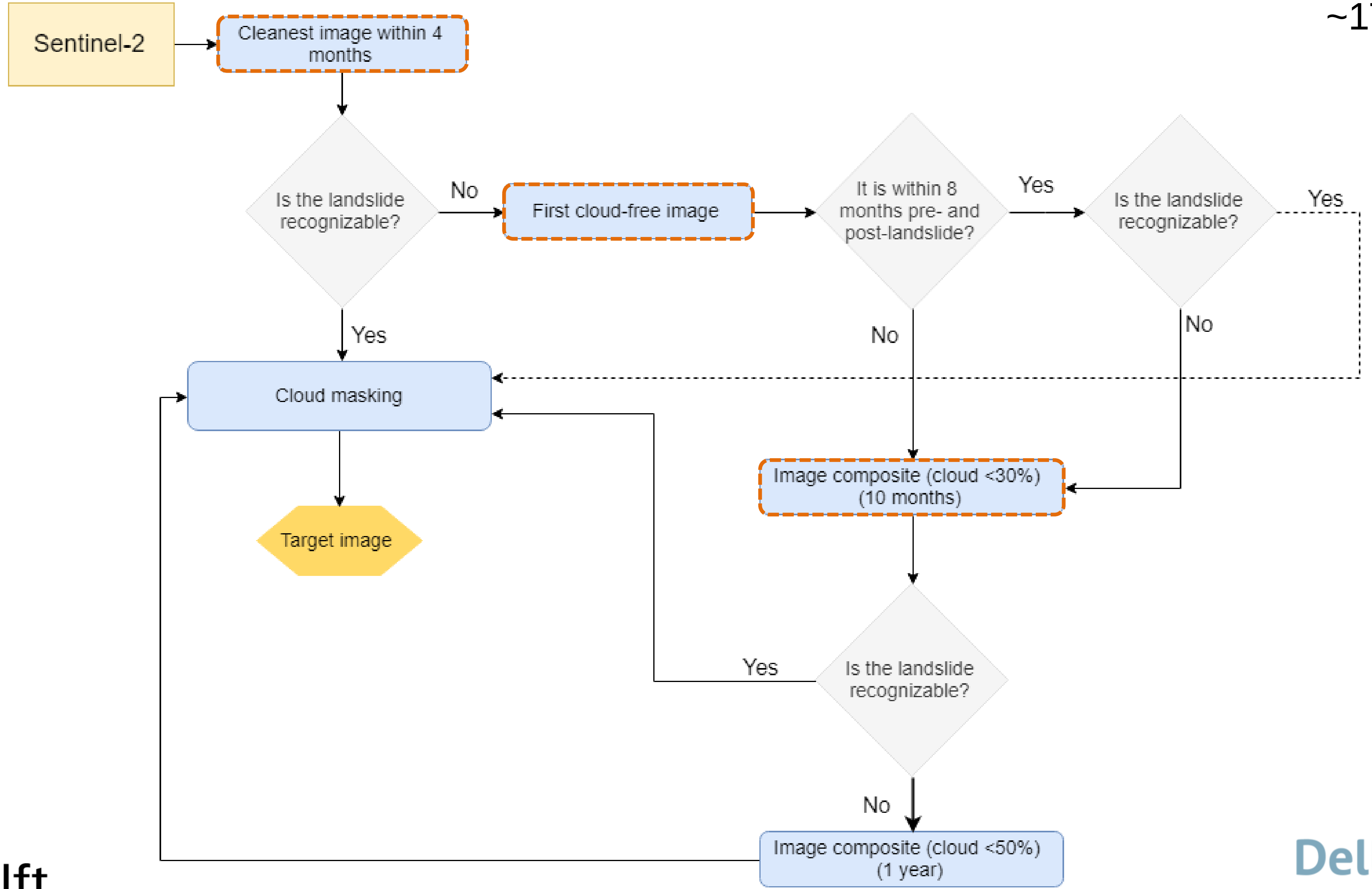
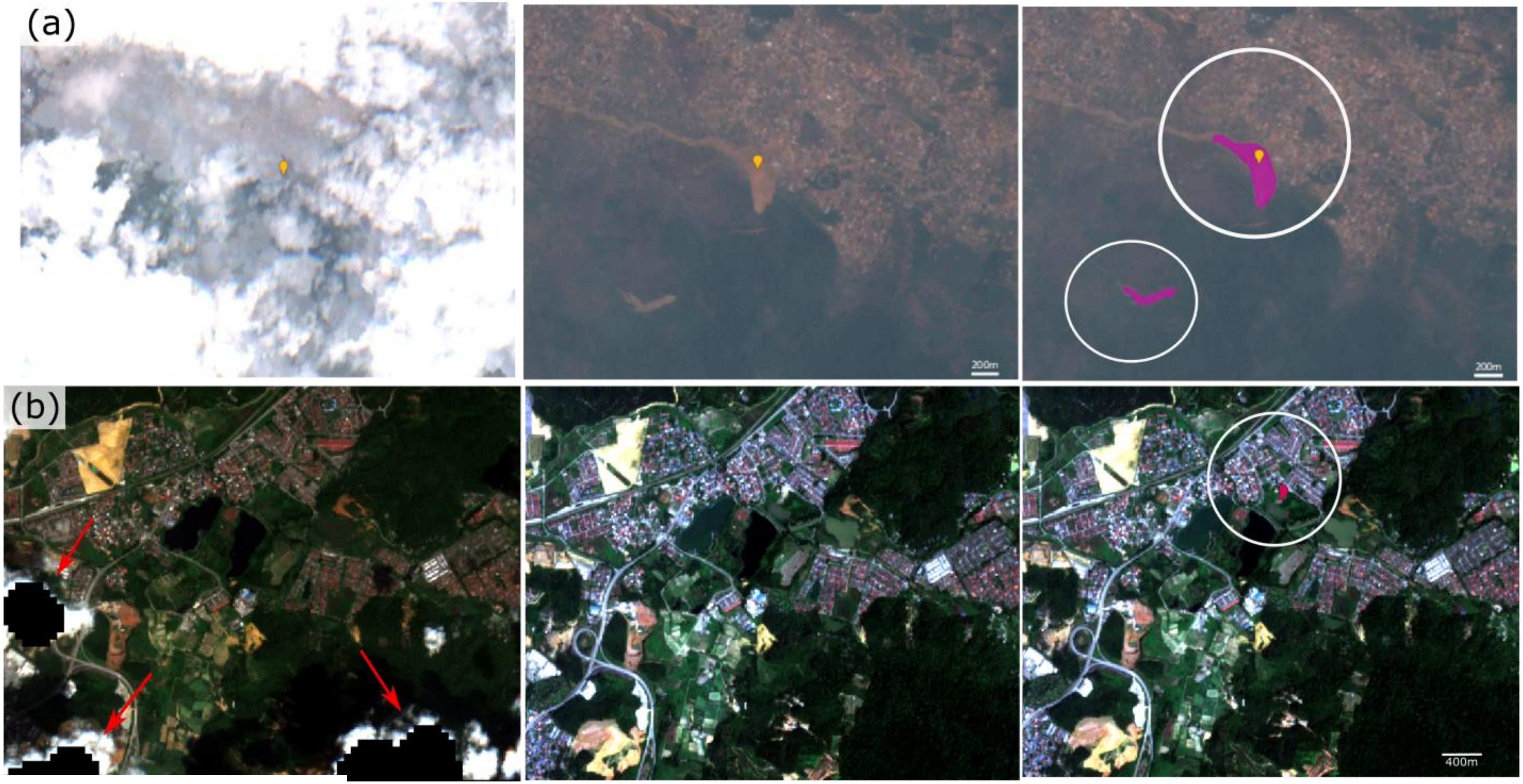


Image set preparation

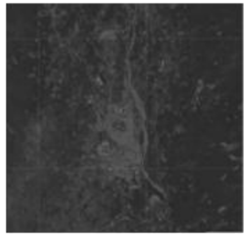


Features computation (RGD)

Cloud-free images

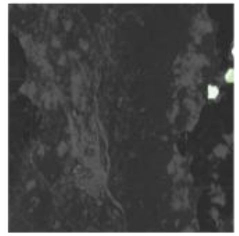
Pre-event Post-event

Band ratioing



Pre-event

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

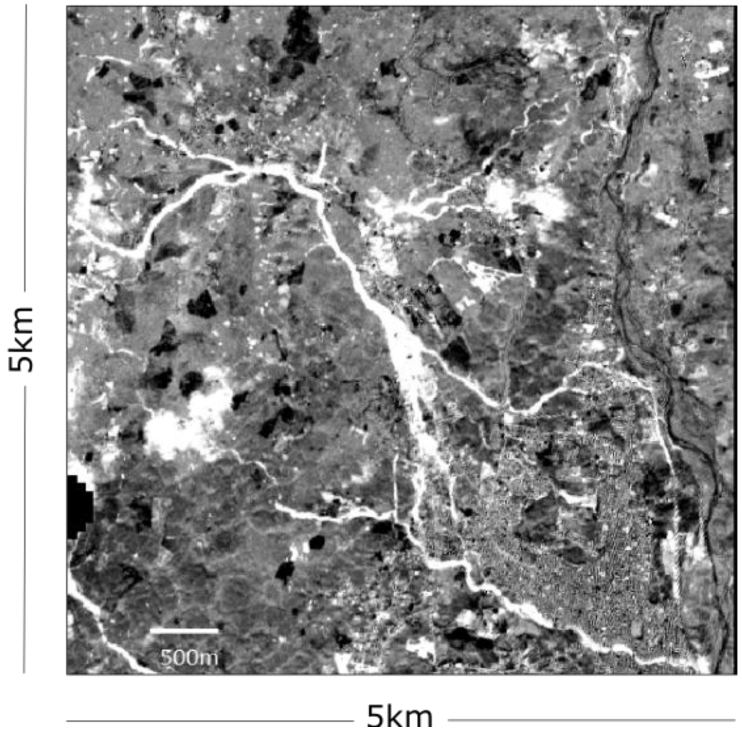


Post-event

Image difference

$$rgd(i,j) = rg(i,j)_{t_2} - rg(i,j)_{t_1} + c$$

$$(i,j) : i = 1, \ldots, n;\; j = 1, \ldots, m$$



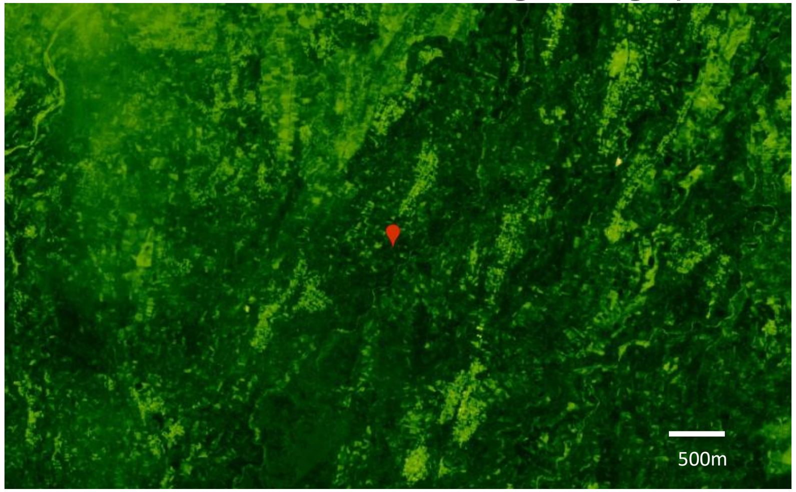
Feature:

Red/Green Difference (RGD)

Features computation (VID)

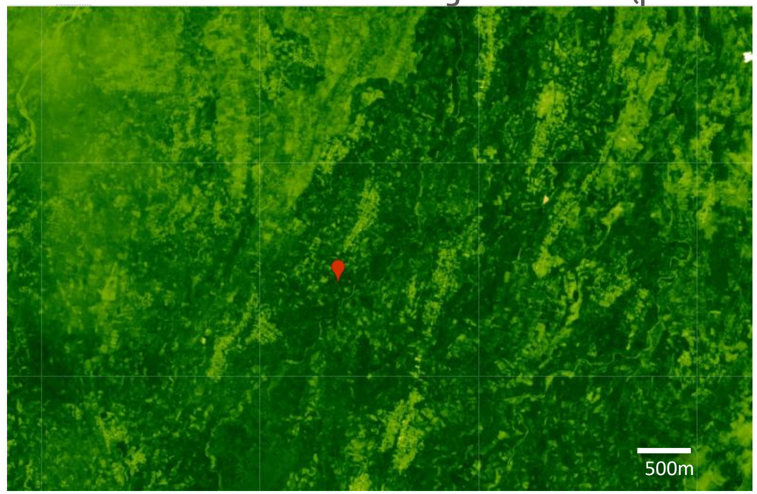
Image normalization

Original image (pre-event)



normalization

Image normalized (pre-event)



Feature:

Vegetation index difference (VID)

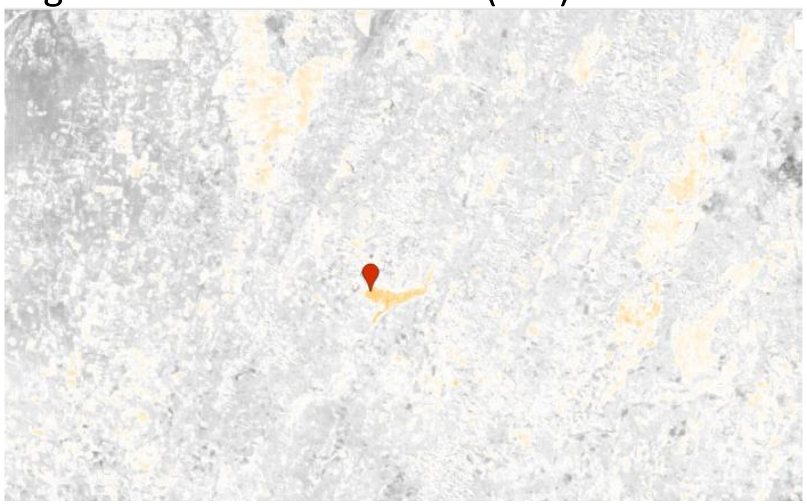
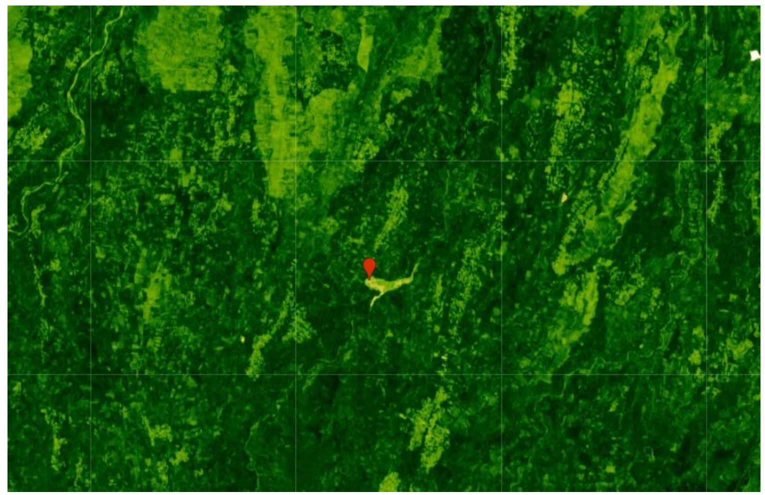


Image difference

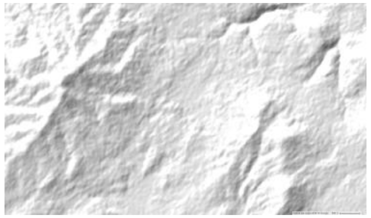
Original image (post-event)



Methodology overview

2015-present

Sentinel-2 (10m)



2006-2011

DEM ALOS (30m)

Free-of-charge datasets

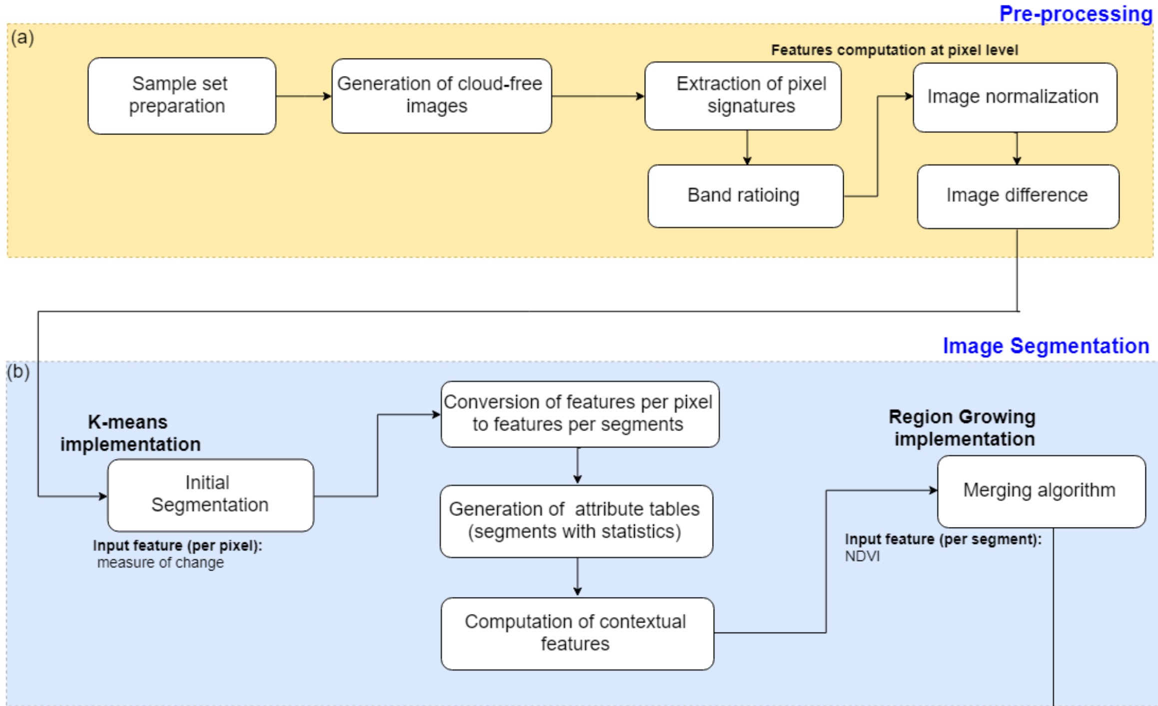


Image classification

Initial Segmentation

Feature: Red/Green Difference (RGD)

Segmentation algorithm: k-means Implementation (Shepherd et al. 2019)

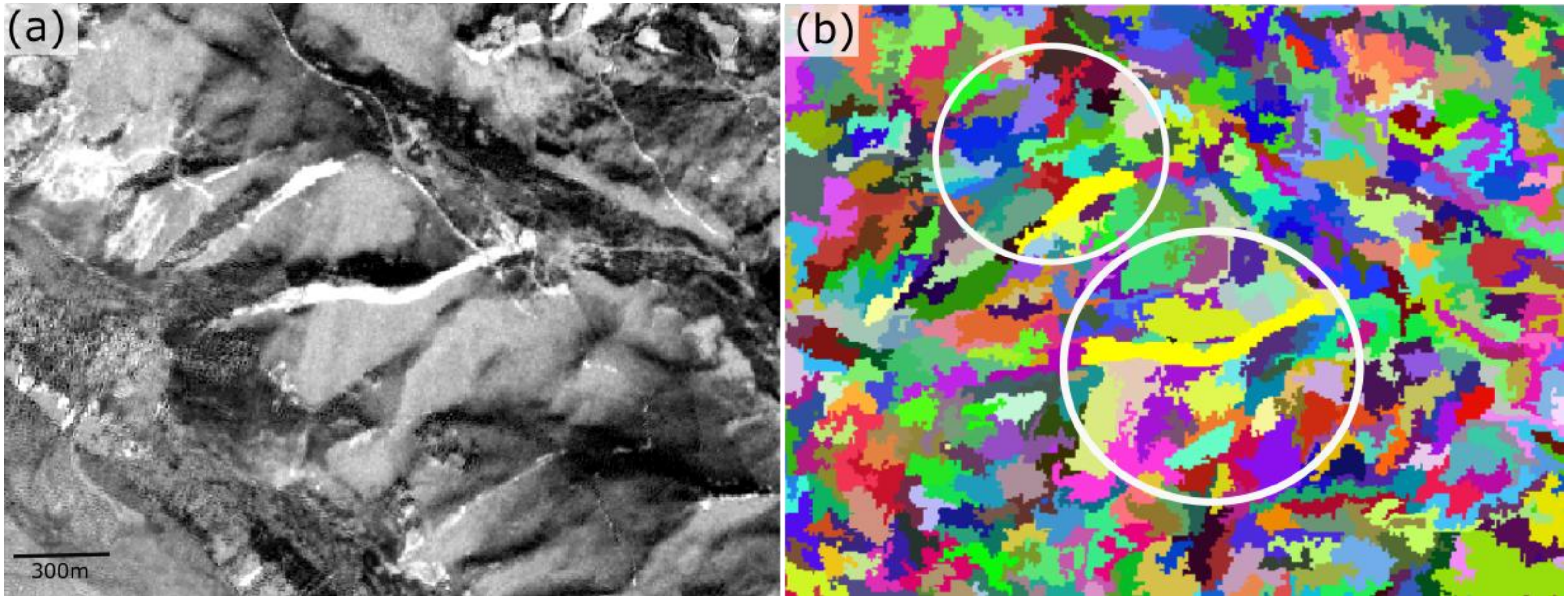


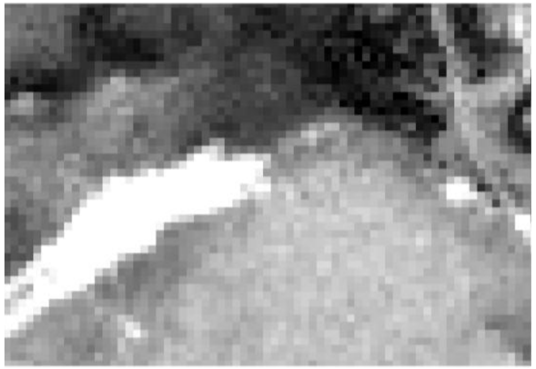
Image difference

Image segmentation

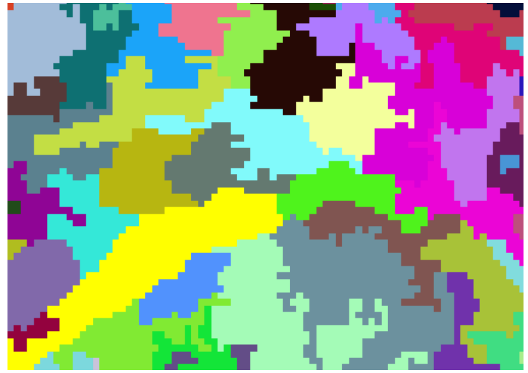
To ensure segmentation of smallest landslides (~100m x 100m)

- Initial K estimated using the Elbow method (k = 8)
- A unique K (for all images) adjusted to a higher value (K=19)
- Minimum number of pixels is fixed to 80*px*

Pixels to Segments



Pixels



Segments

		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

Features at segment level

Segment level

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

Final landslides diagnostic features

TU Delft

Deltares

Initial Segmentation

Challenge → multi-scale objects

- ✓ Over-segmentation of *non- landslides* cases
- ✓ Extremely imbalanced dataset
- ✓ Classifier biased toward the majority class
- ✓ Poor performance for the landslide case

Merging algorithm

Merging algorithm: region growing at segment level

Key Considerations:

- ✓ Criteria for **seed** selection
- ✓ **Feature** to define homogeneity
- ✓ **Threshold** of homogeneity

Homogeneity criterion

Feature: NDVI

$$Sm = \mid \overline{ndvi}_{ws} - ndvi_{nb} \mid$$

If $Sm < t$, segments are merged

Post-event image

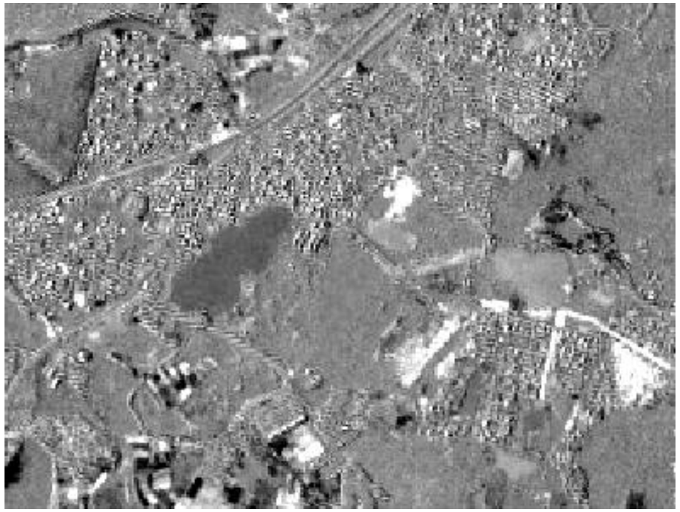
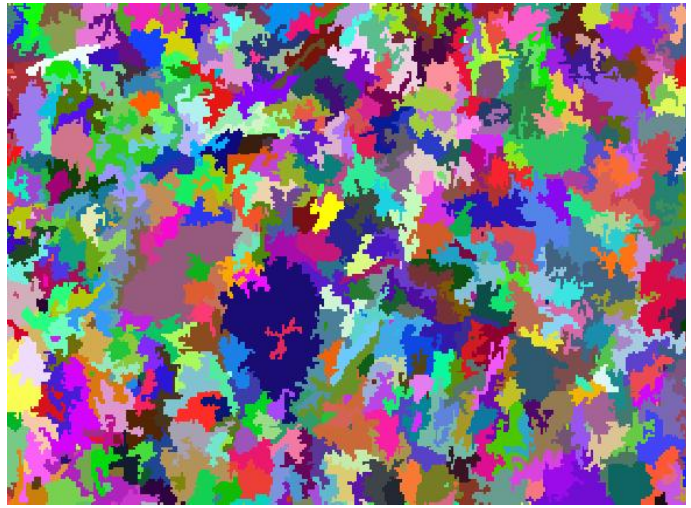


Image difference



Initial segmentation

Homogeneity criterion

Challenge

Different

Images

Types of landslides

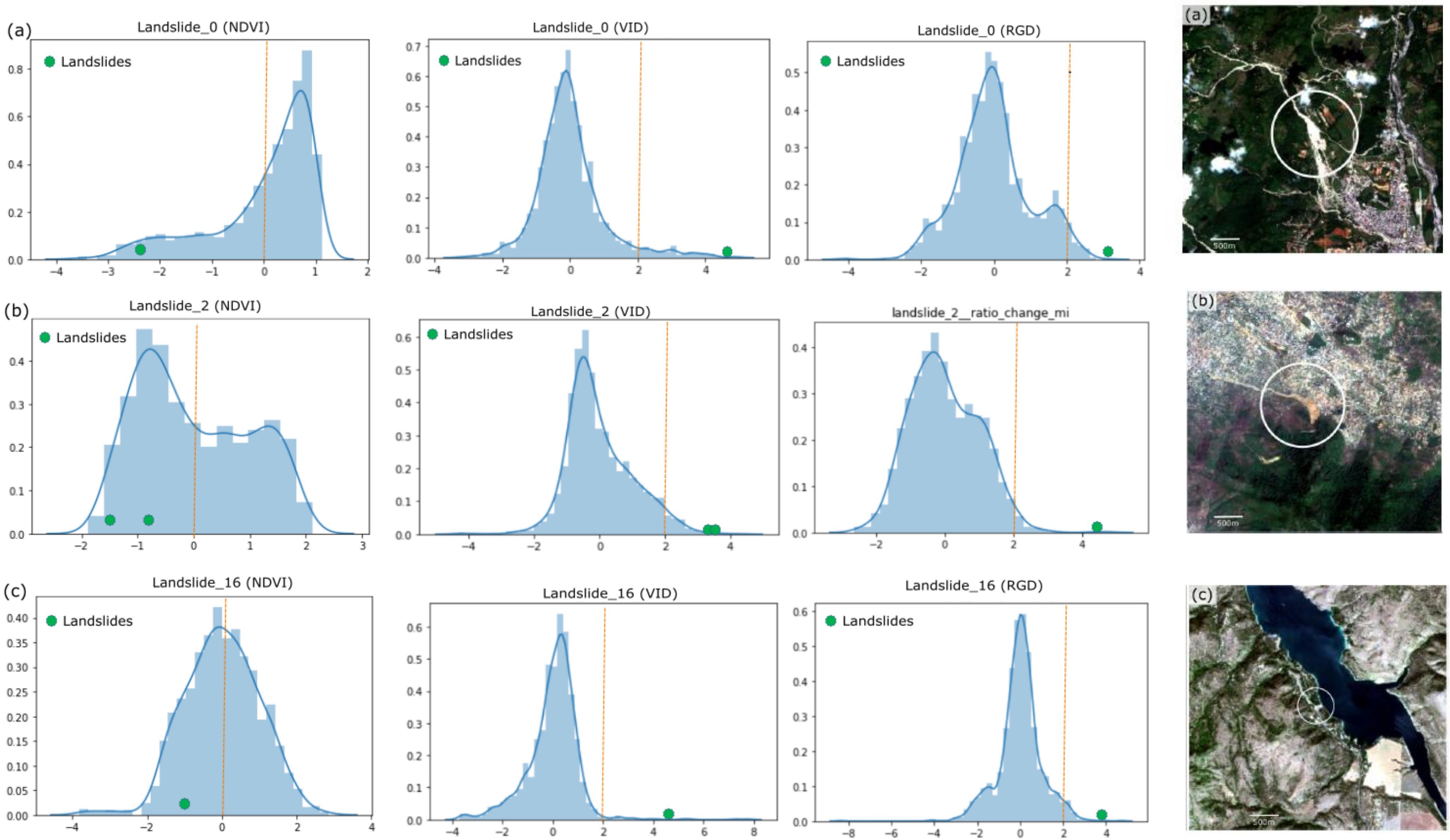
Geomorphological settings

Seasons

Atmospheric conditions

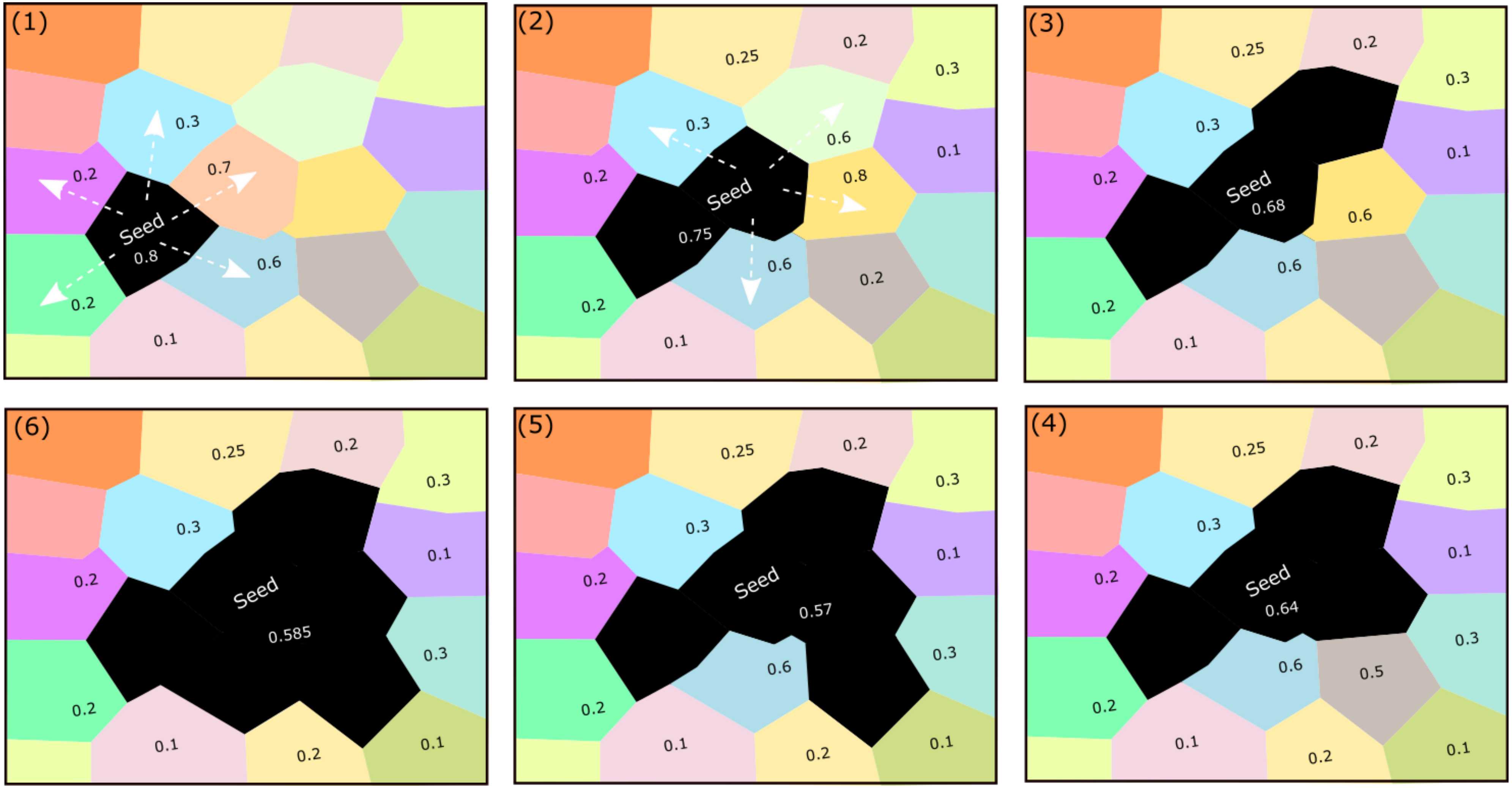
How to define a threshold of similarity for all images ?

Homogeneity criterion



Region growing

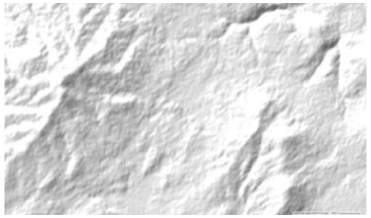
Input Data: List of the Segments
Output Data: List of Regions



Methodology overview

2015-present

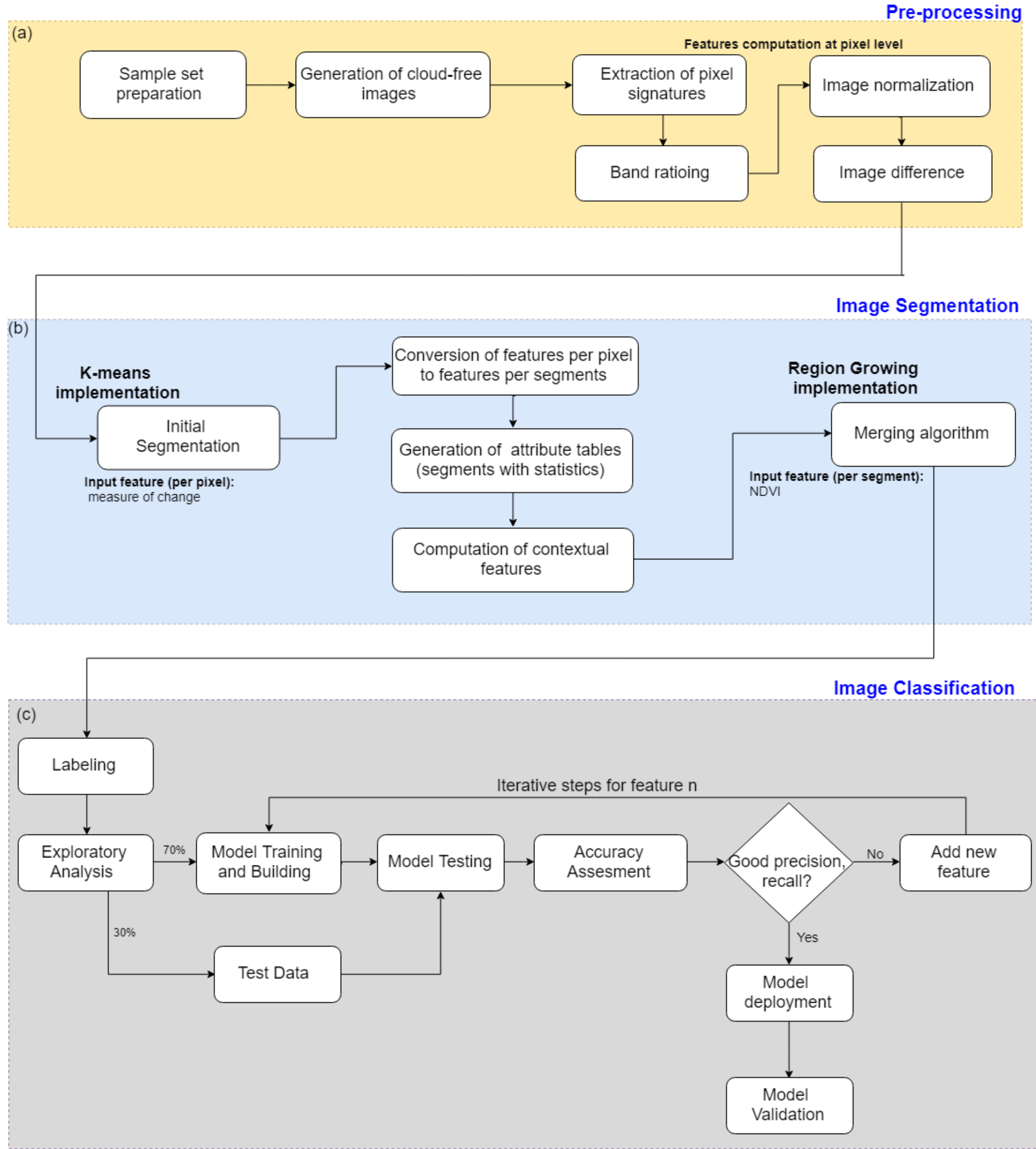
Sentinel-2 (10m)



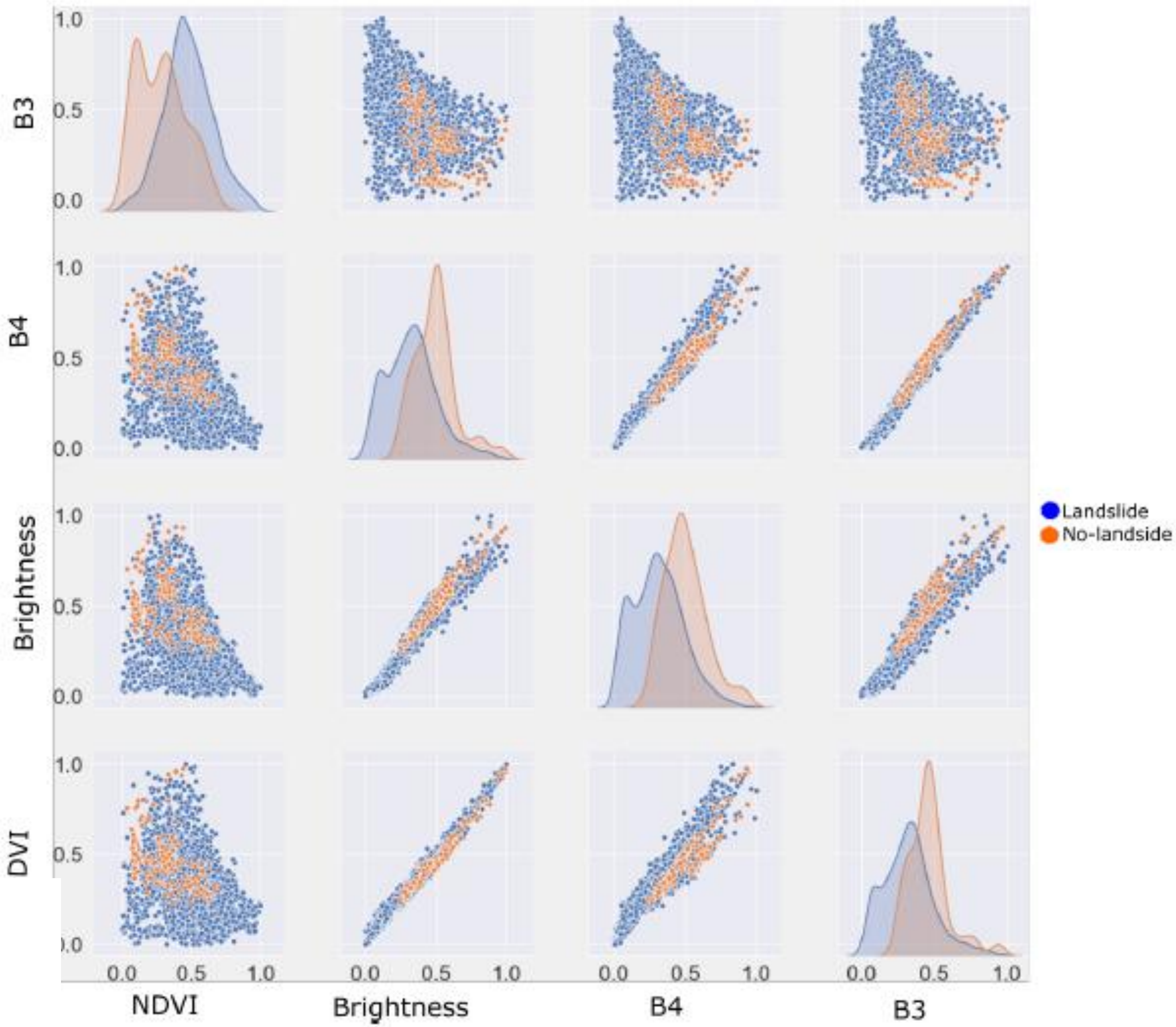
2006-2011

DEM ALOS (30m)

Free-of-charge datasets



Exploratory analysis



Model training and building

Landslide diagnostic features

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

Initial ranking of features

Outline

- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

TU Delft

Deltares

Segments optimization

Coarse threshold	Segments	Reduction	missed	Error
0.1	7296	82%	11	5%
0.2	2749	93.3%	16	8%
0.3	1653	96.3%	24	13%

+ fine threshold = 0.05

K-means

non-landslides segments: 56,563

Ratio 1:225

Merging algorithm

non-landslides segments: 2,749

93.3 %

Ratio 1:14

Model training and building

Total segments = 2,905

landslides = 199

non-landslides = 2,706

Ratio 1:14

Training (70%)

- ***non-landslides*** **=** 1,899
- ***landslides*** **=** 134

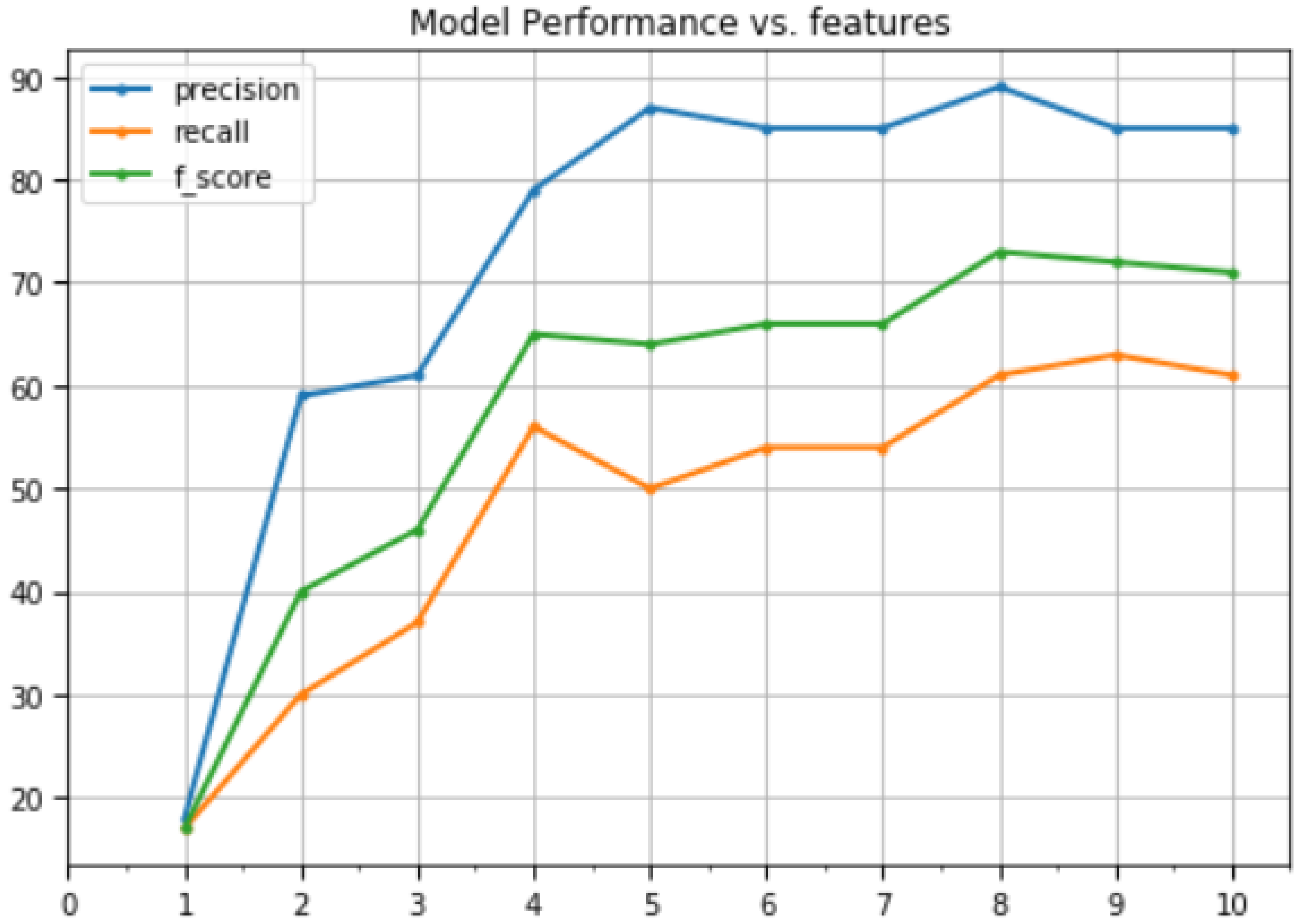
Testing (30%)

- ***non-landslides*** **=** 807
- ***landslides*** **=** 65

TU Delft

Deltares

Model performance

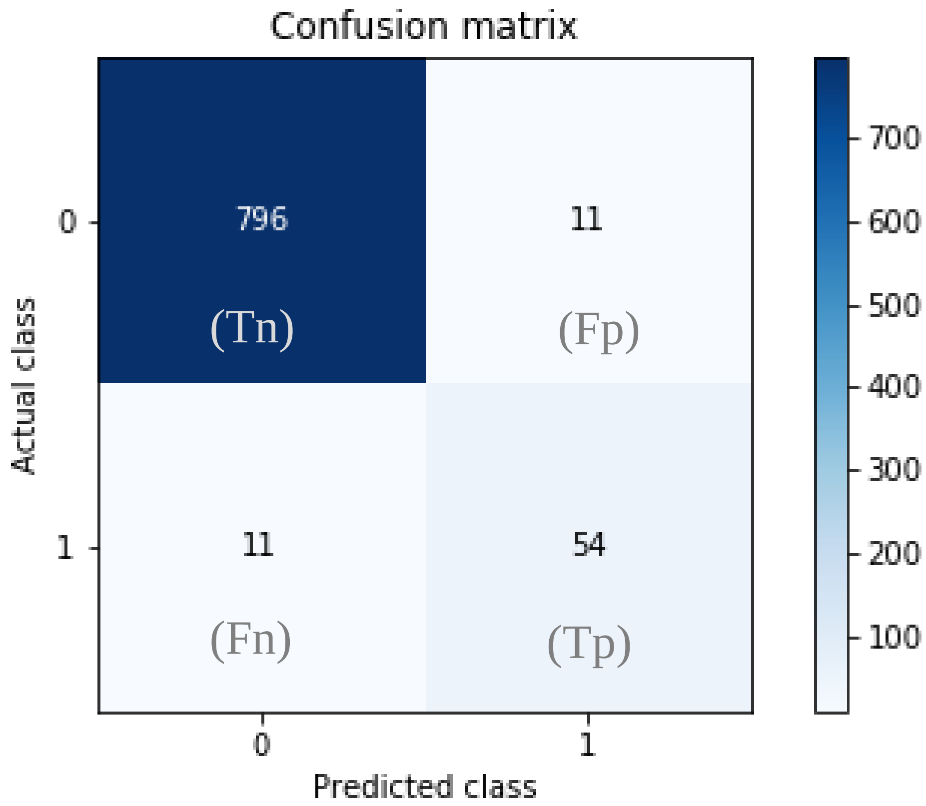


Ranking	Feature
1	NDVI
2	$RGD_{deviation}$
3	$BrightnessD_{deviation}$
4	$VID_{deviation}$
5	Brightness
6	Slope_mean
7	$GNDVI_{deviation}$
8	Slope_max
9	$NDVI_{texture}$
10	Relative_relief

Model performance

landslide = 1

non-landslide = 0



Target class*: landslide*

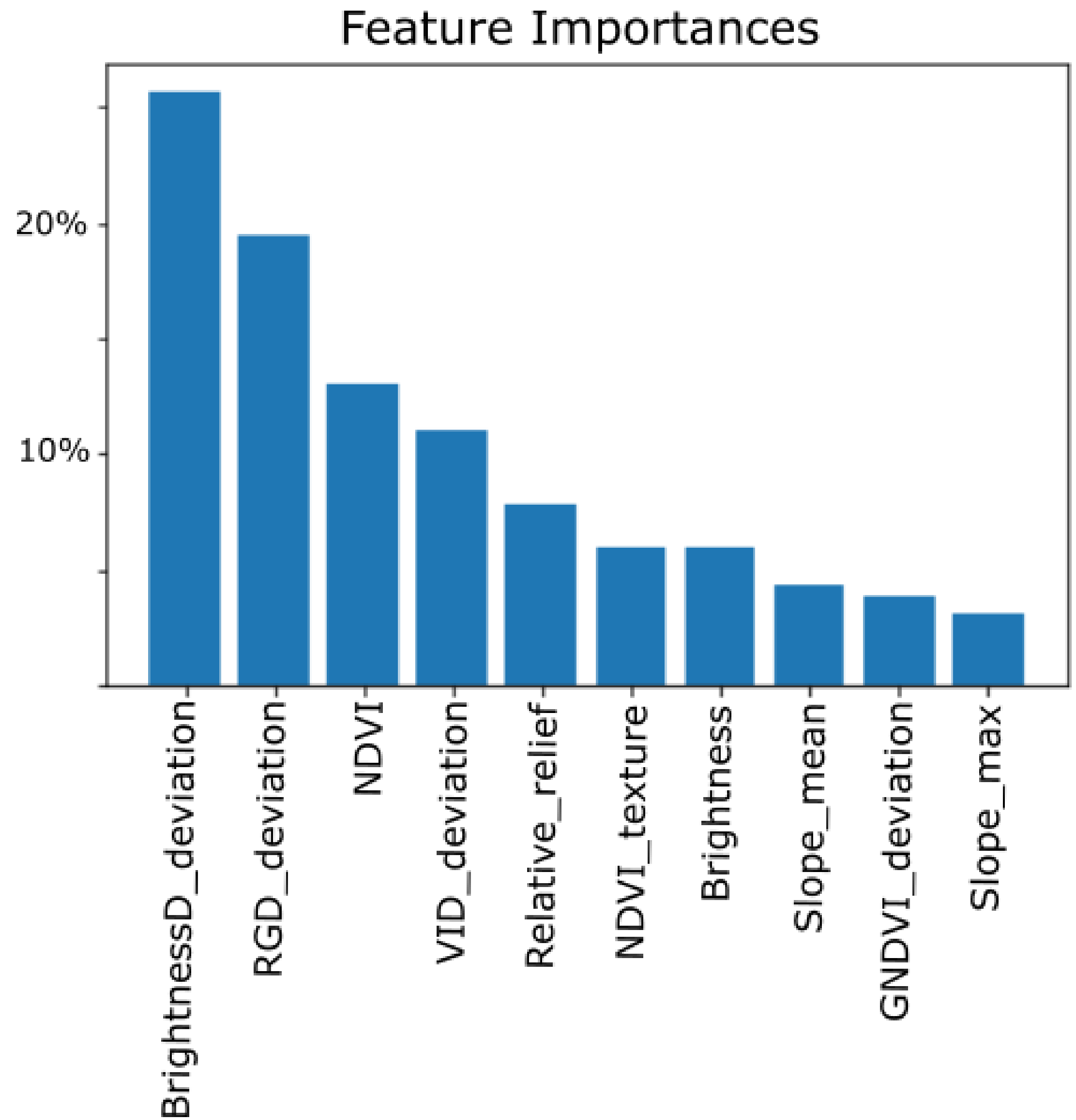
$$\mathcal{P}recision = \frac{\mathcal{T}_p}{\mathcal{T}_p + F_p} = 0.83$$

$$\mathcal{R}ecall = \frac{\mathcal{T}_p}{\mathcal{T}_p + F_n} = 0.83$$

$$\mathcal{F}1 = 2.\frac{precision.recall}{precision + recall} = 0.83$$

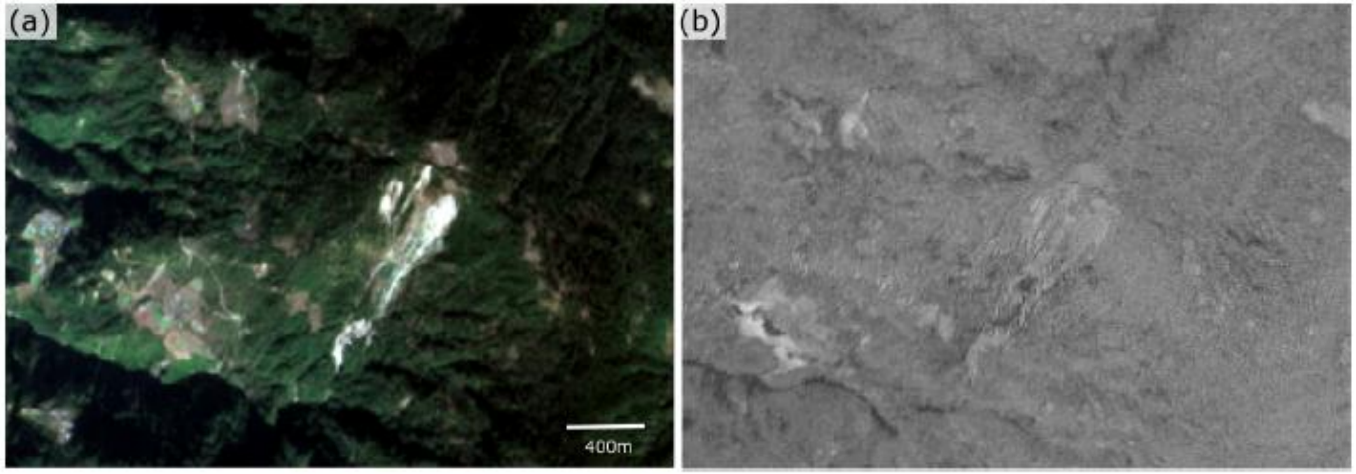
Model performance

Final ranking of features

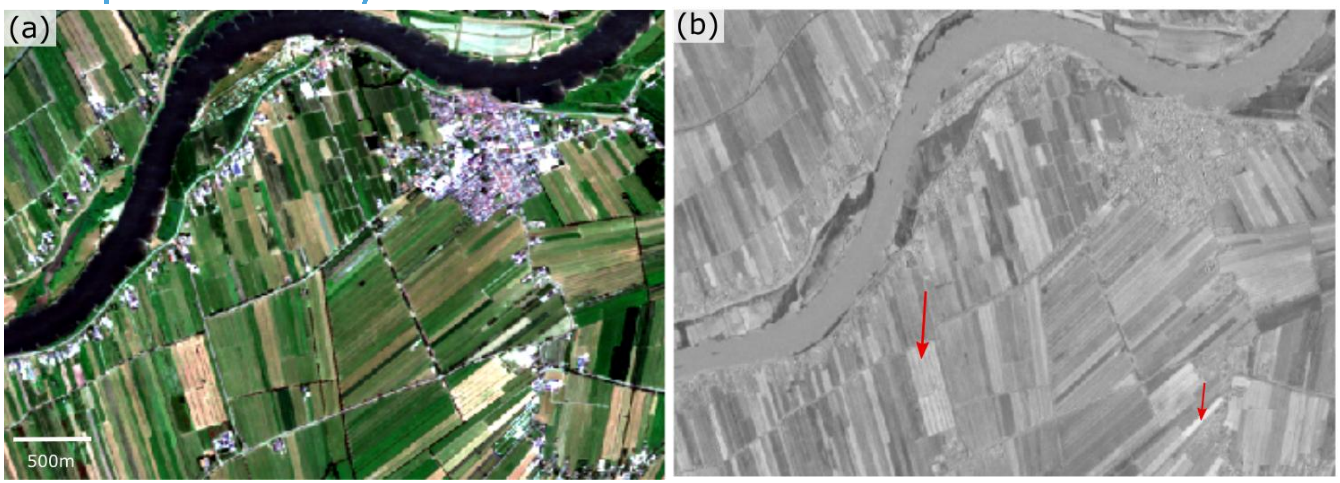


Model validation

(1) Ancient landslides

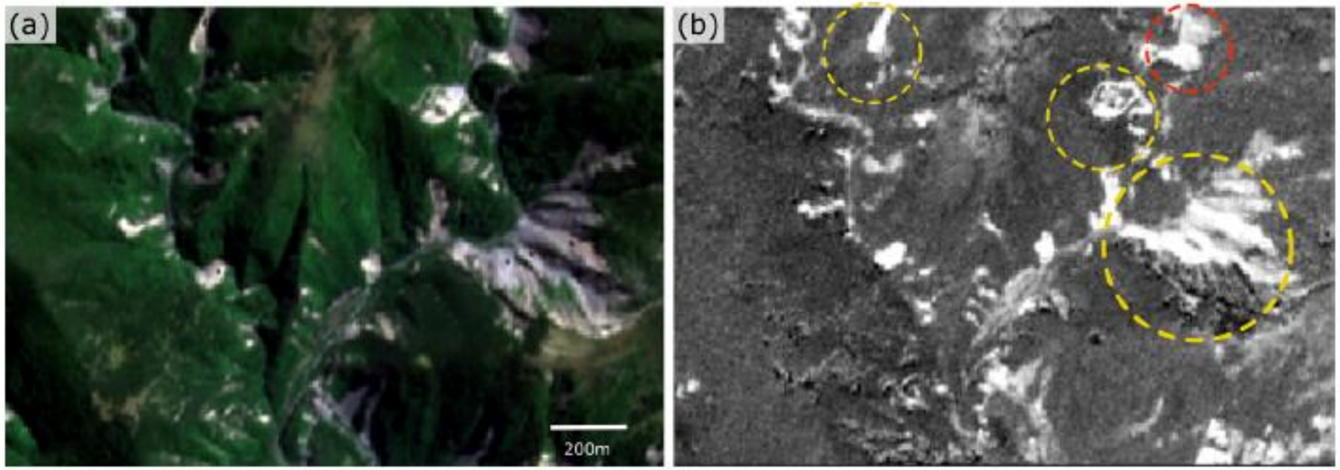


(2) Area with non probability of landslides



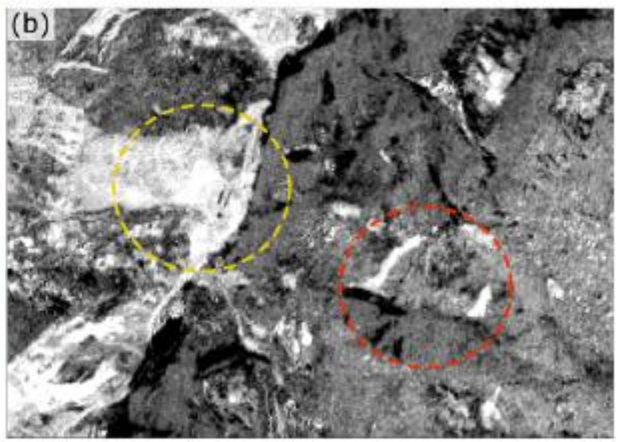
Model validation

(3) Fresh landslides



(4) Nepal case study

NASA tool



Our tool

Out tool

Outline

- Introduction
- Related works
- Methodology
- Results and discussions
- Conclusions and future works

TU Delft

Deltares

Conclusions

How to detect and locate landslides using Machine Learning?

- To what extent can landslides be detected using Sentinel-2 in combination with DEM?

Landslides can be detected using features derived from the spectral information of Sentinel-2 and topographic features from global DEM.

Conclusions

- What are the relevant landslide diagnostic features?

1.	$\mathrm{BrightnessD}_{deviation}$	26%
2.	$\mathrm{RGD}_{deviation}$	19%
3.	NDVI	13%
4.	$\mathrm{VID}_{deviation}$	12%
5…10	Rest of the features	<10% each

Conclusions

- What is the best segmentation strategy?

Initial Segmentation (Red/Green difference) → Detailed segmentation of landslides

Merging algorithm (NDVI) → To balance the dataset (reduce the number of non-landslides cases)

Conclusions

- How to exploit features per pixels to produce features per segments?

  ✓ Features at pixel level are grouped into segments

  ✓ Statistical measures: mean and maximum

Mean	➡	all features
Maximum	➡	slope

Conclusions

- What is the most appropriate Machine Learning technique?

  Random Forest

  ✓ Non-parametric
  ✓ Best performance in OBIA
  ✓ Non-complex implementation
  ✓ Few tuning parameters
  ✓ Can handle imbalanced datasets

Conclusions

- What is the accuracy of most appropriate Machine Learning technique?

✓ Precision = 83%

✓ Recall= 83%

✓ F1-score= 83%

Conclusions

- First attempt of a general method to detect landslides
- Semi-automatic method
- Applicable and re-usable
- Able to work with mixed-landcover (urban-vegetated)
- Limitations in areas with perennial snow, high sedimentation rates, regions with non-vegetation.

TU Delft

Deltares

Future works

- Increase the number of training samples
- Remove redundant information
- Apply similar methodology to extract features from SAR images
- Explore an strategy to include automatic derived thresholds

TU Delft

Deltares

Thanks for your attention!

Questions ?

Tools

Storage	Pre-processing	Processing	Visualization
PostgreSQL Google fusion tables	Google Earth Engine	GDAL NumPy RSGISLib	

Source code: https://github.com/mhscience/landslides_detection