Kyoto University

October 4, 2023

RESPONSIBLE AI FOR EARTH OBSERVATION

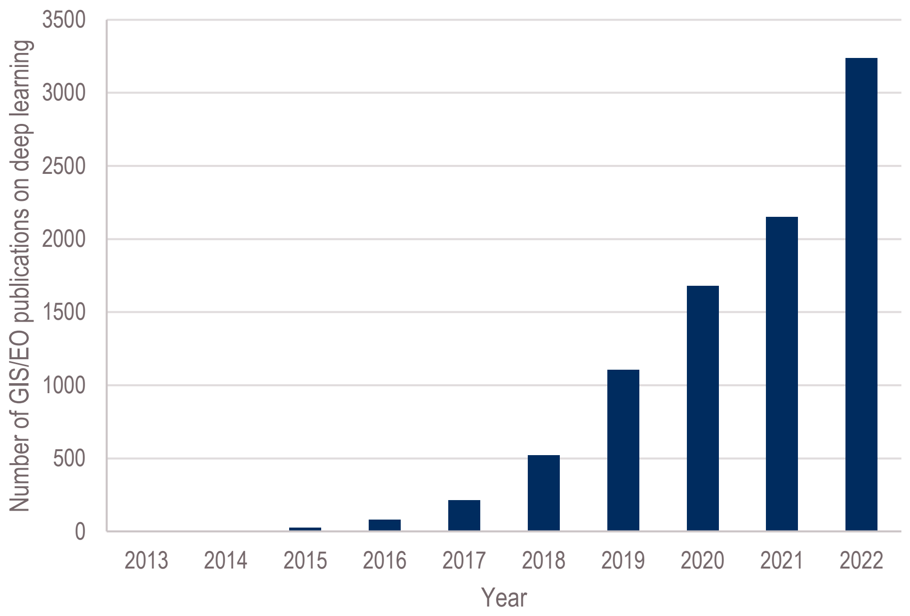
CAROLINE GEVAERT

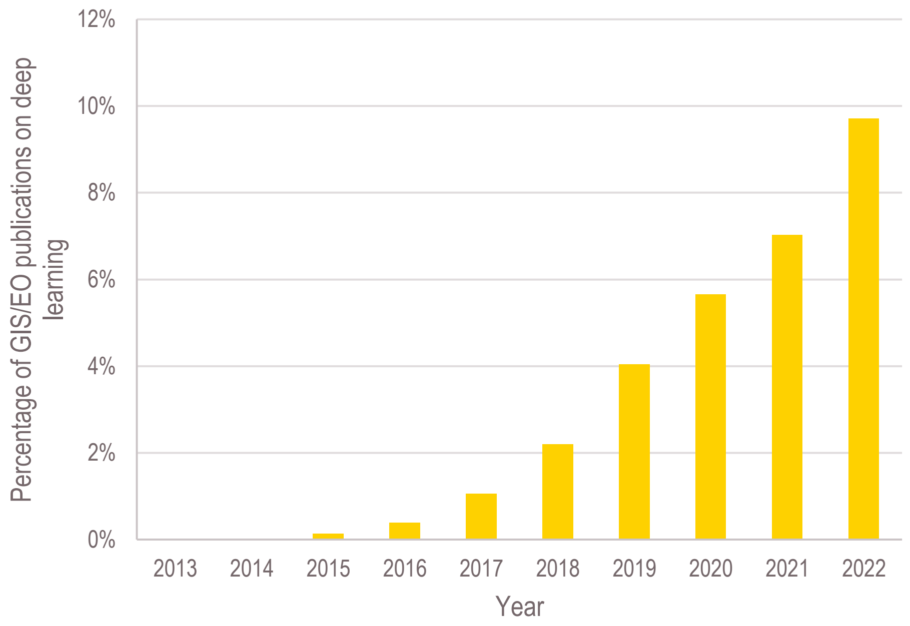
c.m.gevaert@utwente.nl

*This work is partially funded by the NWO-VENI project no. 18091

UNIVERSITY OF TWENTE. ITC

AI IN EARTH OBSERVATION



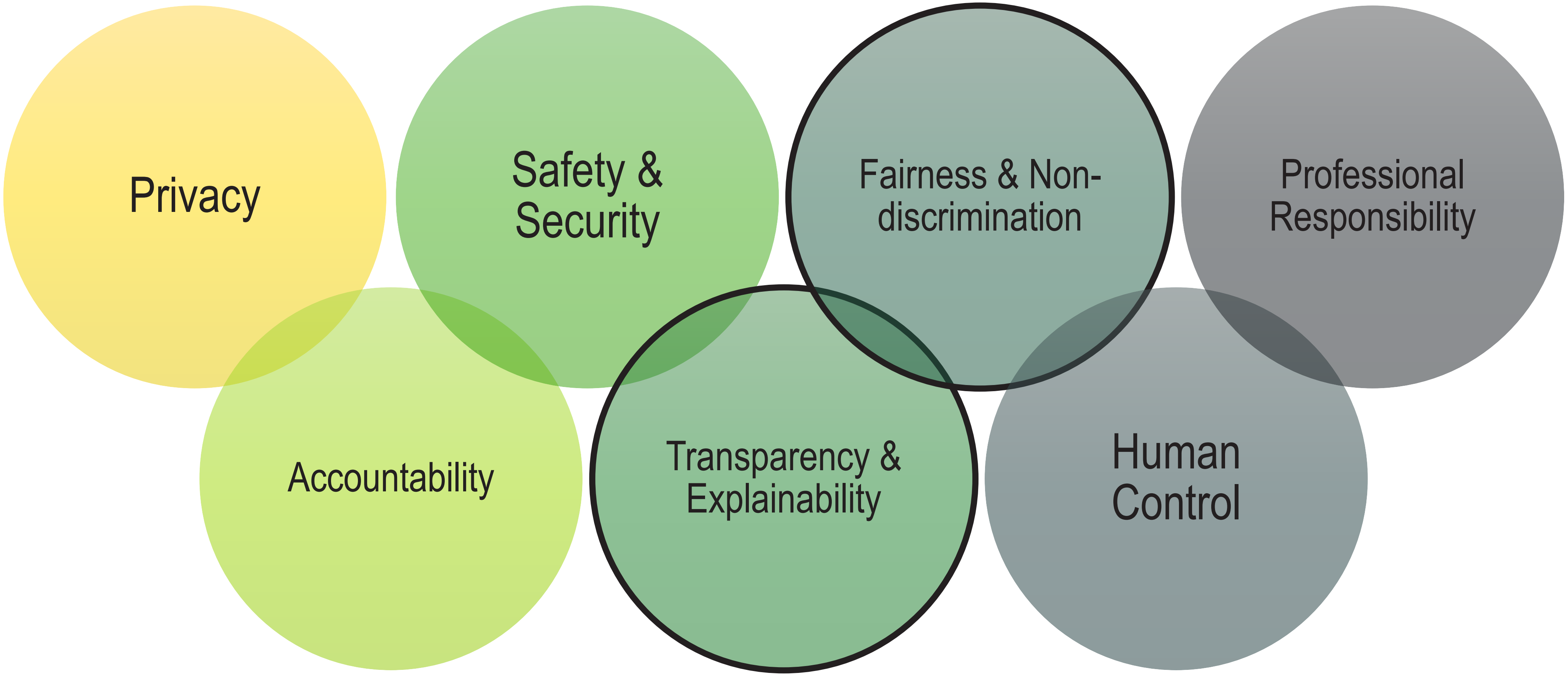


Academic publications in Scopus
TITLE-ABS-KEY ("deep learning" AND ("GIS" OR "Earth Observation" OR "Remote Sensing" OR "geospatial"))

Source: Scopus

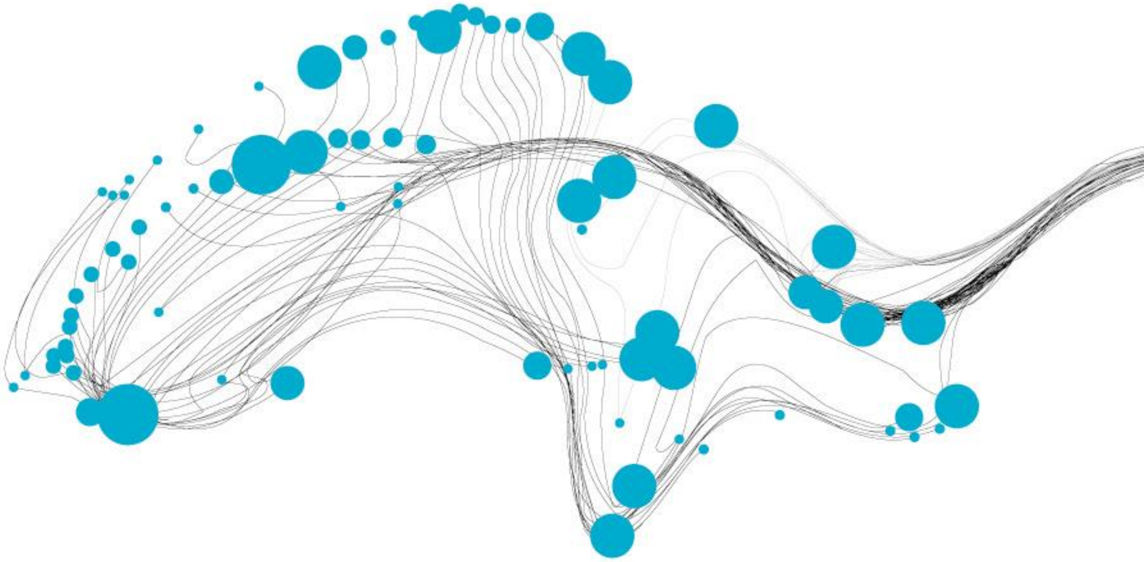
UNIVERSITY OF TWENTE.

GUIDELINES & FRAMEWORKS



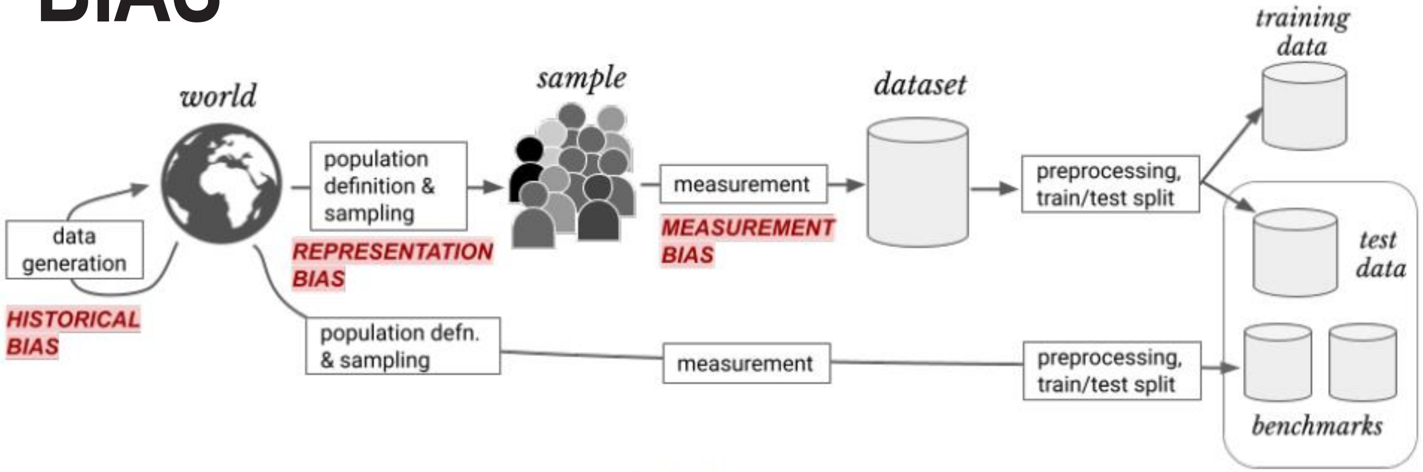
Source: Fjeld et al.(2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No 2020-1

FAIRNESS / BIAS

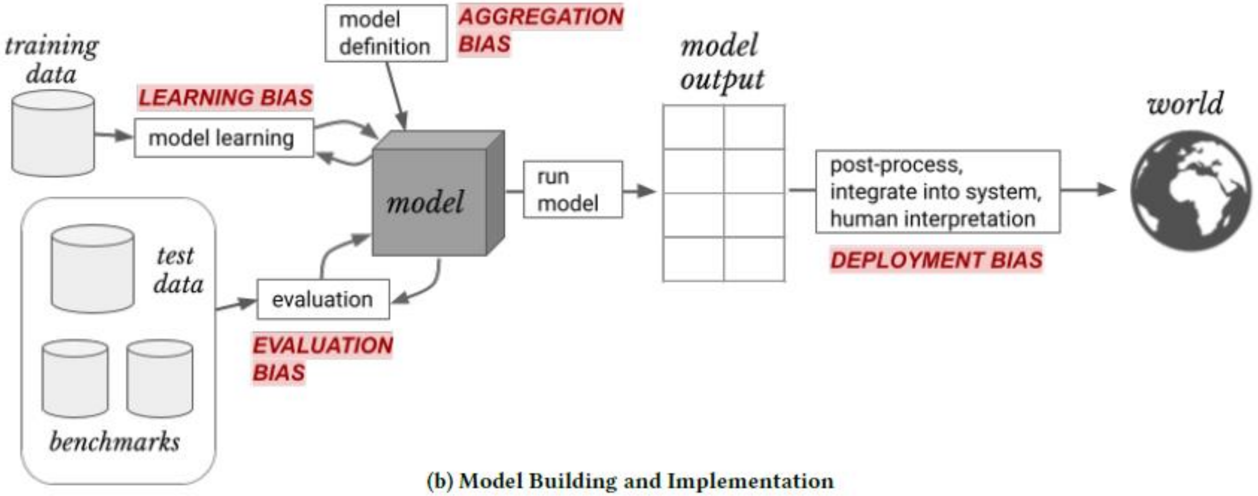


UNIVERSITY
OF TWENTE.

TYPES OF BIAS



(a) Data Generation



(b) Model Building and Implementation

Suresh, H. & Guttag, J. V. A Framework for Understanding Unintended Consequences of *Machine Learning*. (2020).

Source: Gevaert, C. (2022) *EARSeL Cyprus 2022.*

UNIVERSITY OF TWENTE. ITC

FAIRNESS?

	1	2
1	TP	FN
2	FP	TN

SIMPLE

	1	2	3
1	TP		
2		TP	
3			TP

MULTI-CLASS

Attribute 1

	1	2
1	TP	FN
2	FP	TN

Attribute 2

	1	2
1	TP	FN
2	FP	TN

AUDIT FOR BIASES

BIAS METRICS

Table 2: List of group metrics.

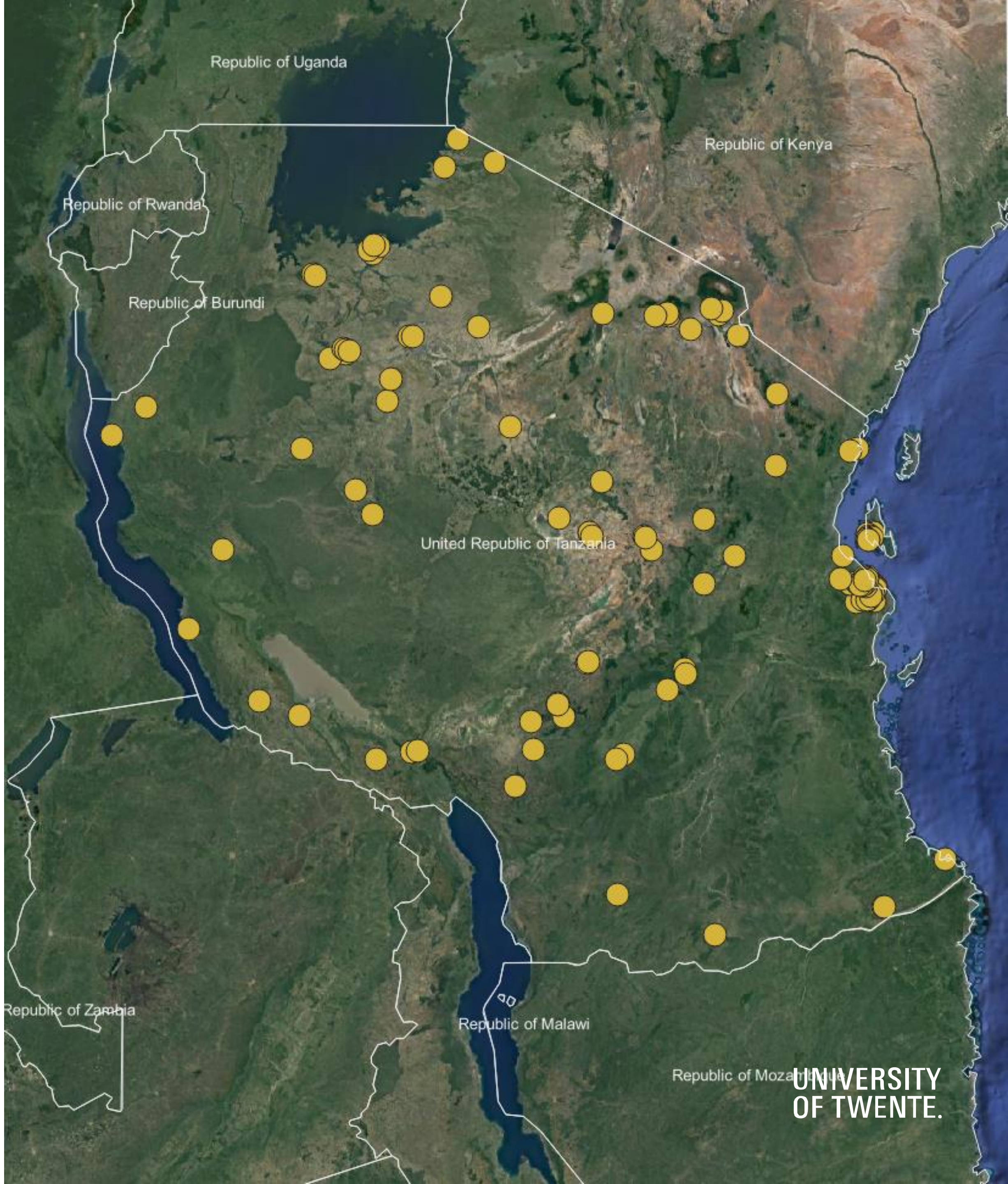
Name	Notation	Definition
Prevalence	$Prev_g = LP_g \ / \ \lvert g \rvert = \Pr(Y{=}1 \lvert A{=}a_i)$	fraction of entities within a group which true outcome was positive.
Predicted Prevalence	$PPrev_g = PP_g \ / \ \lvert g \rvert = \Pr(\widehat{Y}{=}1 \lvert A{=}a_i)$	fraction of entities within a group which were predicted as positive.
Predicted Positive Rate	$PPR_g = PP_g \ / \ K = \Pr(a{=}a_i \lvert \widehat{Y}{=}1)$	fraction of the entities predicted as positive that belong to a certain group.
False Discovery Rate	$FDR_g = FP_g \ / \ PP_g = \Pr(Y{=}0 \lvert \widehat{Y}{=}1, A{=}a_i)$	fraction of false positives of a group within the predicted positive of the group
False Omission Rate	$FOR_g = FN_g \ / \ PN_g = \Pr(Y{=}1 \lvert \widehat{Y}{=}0, A{=}a_i)$	fraction of false negatives of a group within the predicted negative of the group
False Positive Rate	$FPR_g = FP_g \ / \ LN_g = \Pr(\widehat{Y}{=}1 \lvert Y{=}0, A{=}a_i)$	fraction of false positives of a group within the labeled negative of the group
False Negative Rate	$FNR_g = FN_g \ / \ LP_g = \Pr(\widehat{Y}{=}0 \lvert Y{=}1, A{=}a_i)$	fraction of false negatives of a group within the labeled positives of the group

Source: Aequitas documentation
https://dssg.github.io/aequitas/metrics.html

EXAMPLE – BUILDING DETECTION IN TANZANIA

- 100 study areas of 250 x 250 m
- Distributed over poverty groups, city size
- Manually digitized over Google Satellite imagery
  → could cause issues regarding temporal differences or shift in imagery

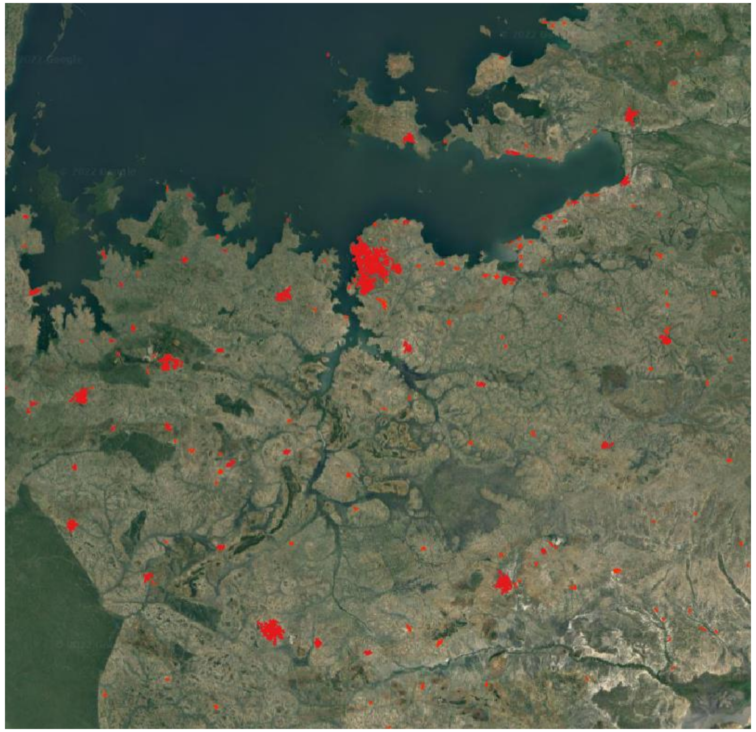
Source: Gevaert, C. (2022) *EARSeL Cyprus 2022.*



UNIVERSITY OF TWENTE.

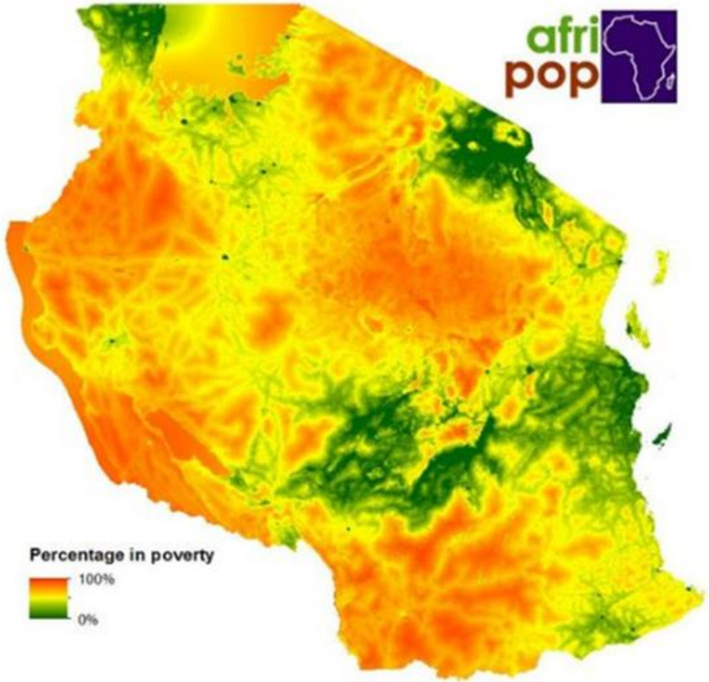
SENSITIVE ATTRIBUTES

CITY SIZE



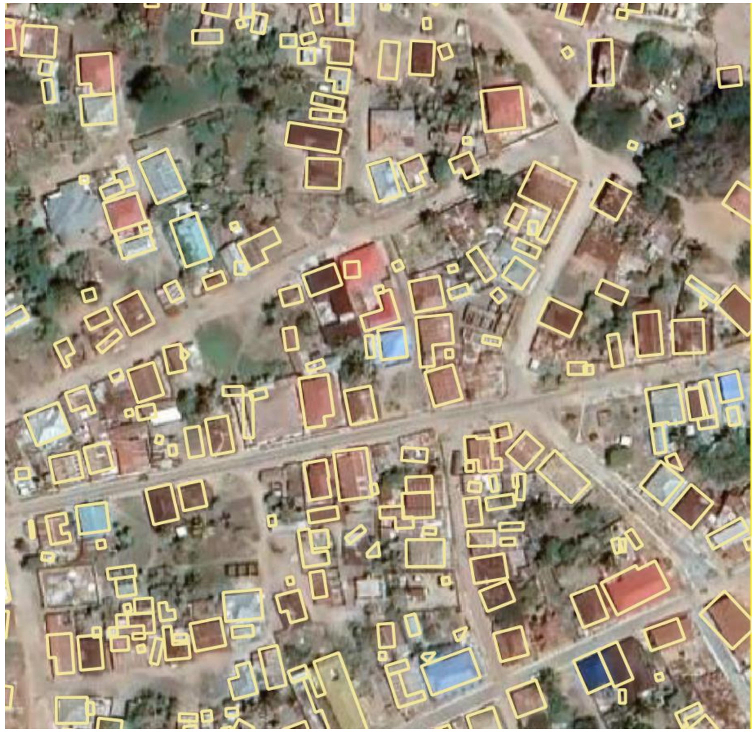
WSF-2019
https://download.geoservice.dlr.de/WSF2019/files/

POVERTY



WorldPop
DOI : 10.5258/SOTON/WP00290

BUILDING SIZE



ACCURACY – EXAMPLE LOW POVERTY, BIG CITY

OSM BING GOOGLE

Source: Gevaert, C. (2022) *EARSeL Cyprus 2022.*

UNIVERSITY OF TWENTE.

ACCURACY – EXAMPLE HIGH POVERTY, RURAL

OSM

BING

GOOGLE

Source: Gevaert, C. (2022) *EARSeL Cyprus 2022.*

UNIVERSITY OF TWENTE. | ITC

PRECISION

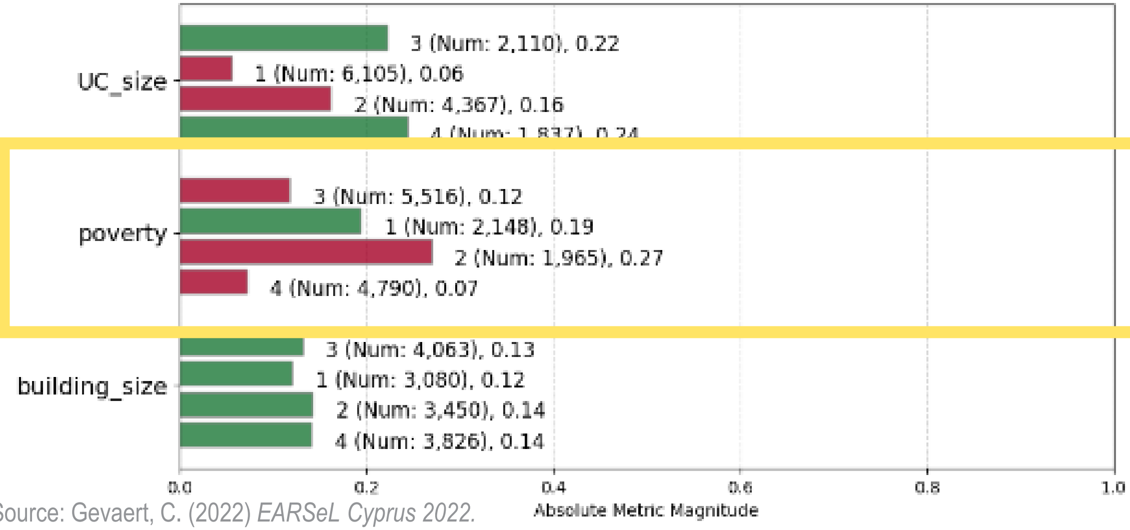
$$Precision_g = \frac{TP_g}{LP_g} = Pr(\widehat{Y} = 1 | Y = 1, A = a_i)$$

Fraction of true positives of a group divided by the labelled positives (=true positive + false positive) of the group.

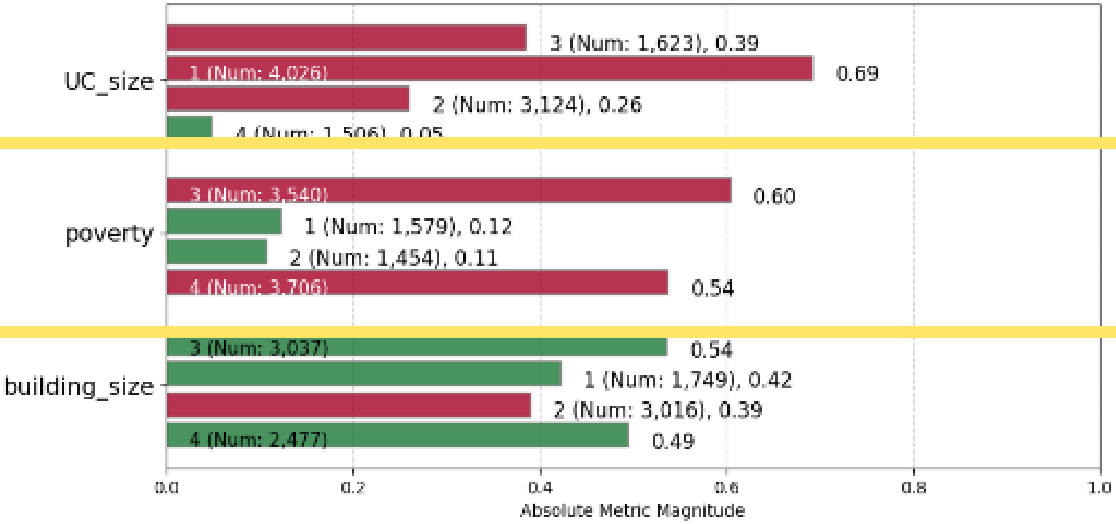
OSM
More precise for larger cities, biased against smaller cities
More precise for lower poverty levels, biased for poverty

Google
More precise for rural areas, biased against smaller cities
More precise for higher poverty levels, biased for poverty

OSM

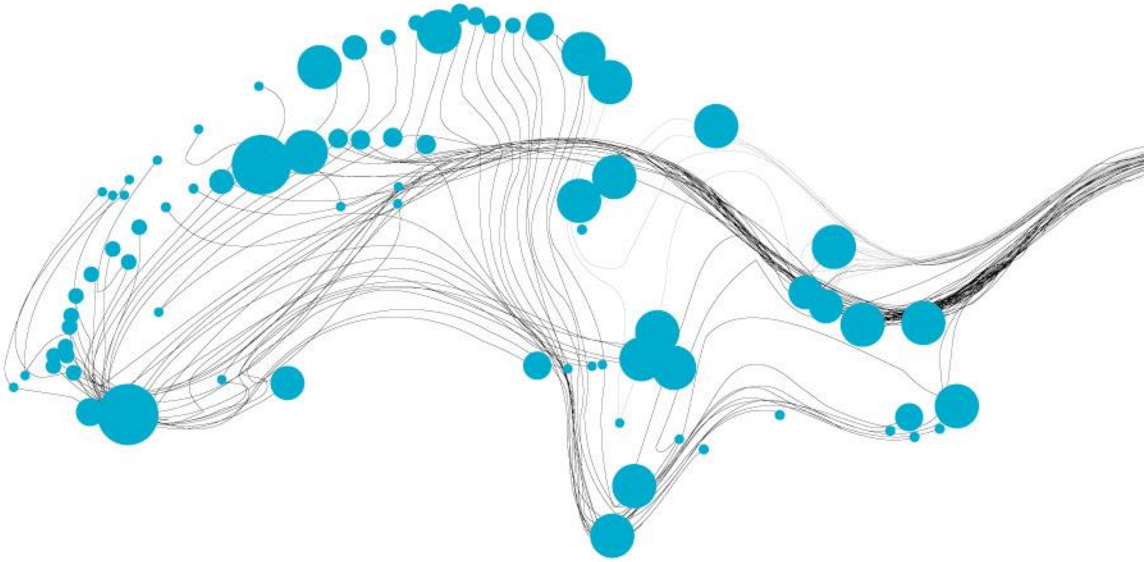


GOOGLE



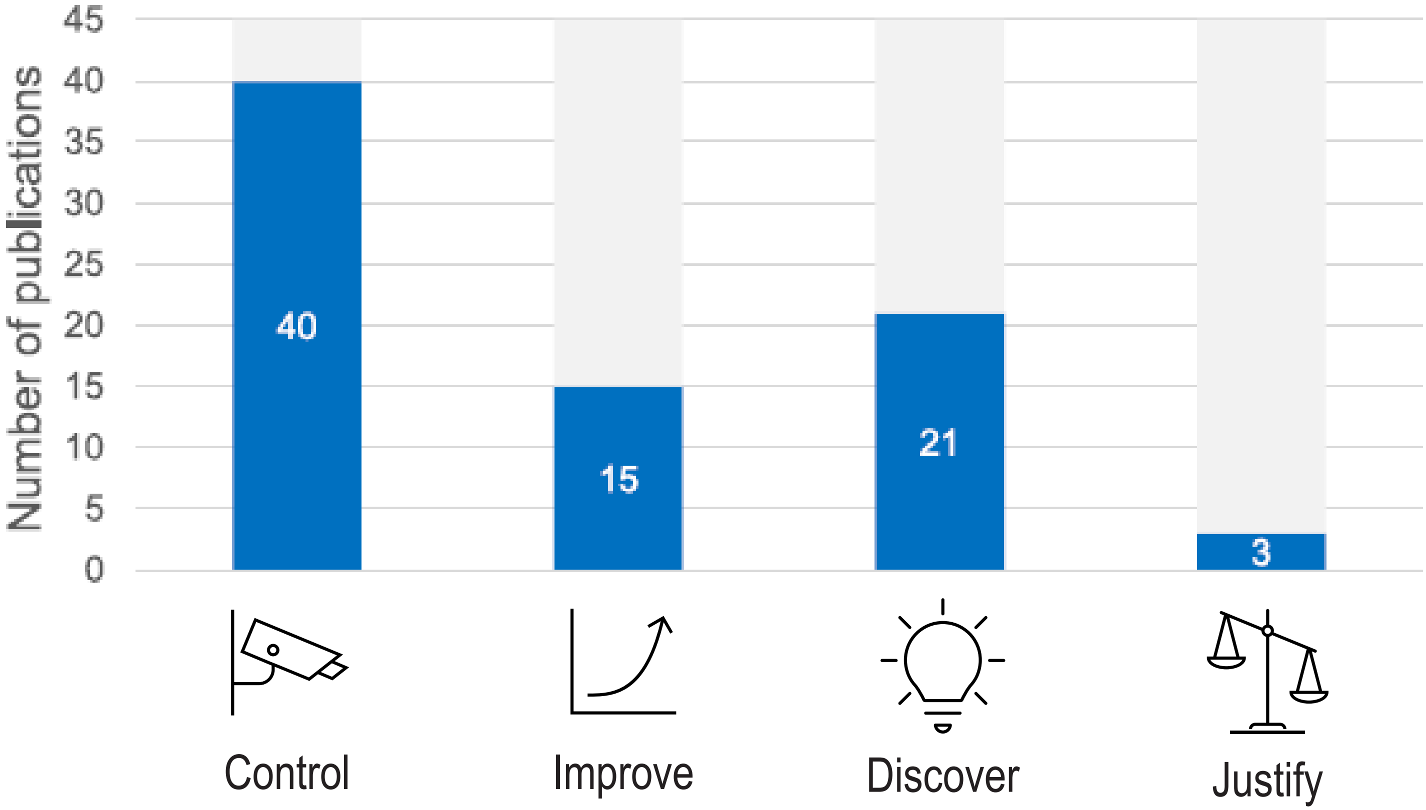
Source: Gevaert, C. (2022) *EARSeL Cyprus 2022.*

EXPLAINABILITY



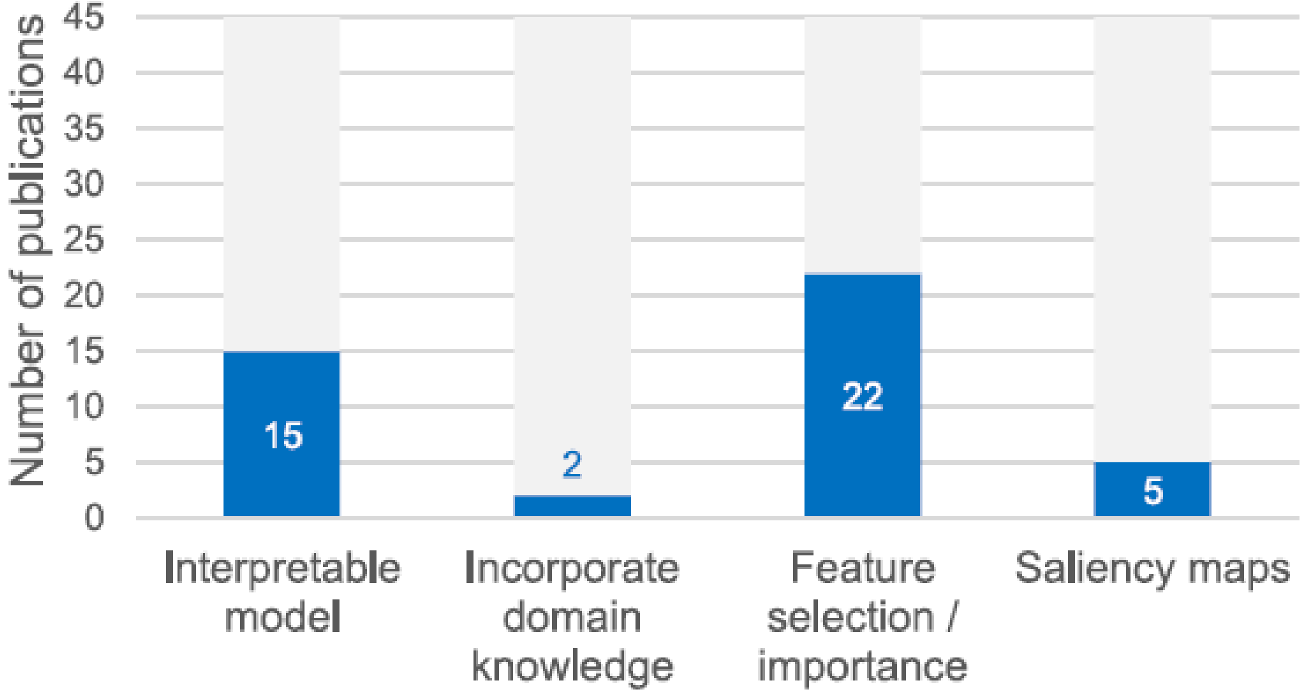
UNIVERSITY
OF TWENTE.

MOTIVATIONS FOR EXPLAINABLE AI



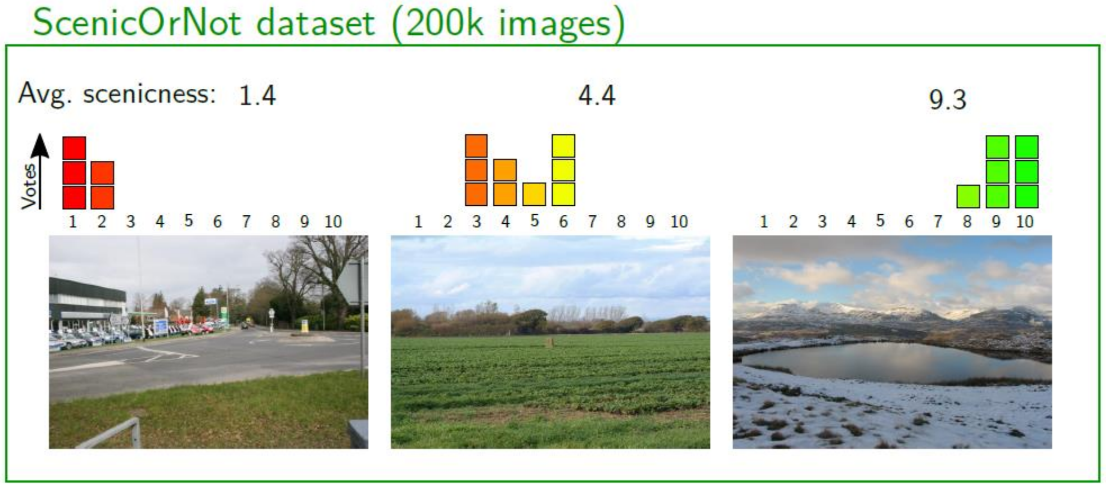
Gevaert (2022) Explainable AI for earth observation: A review including societal and regulatory perspectives,
Int. Journal of Applied EO and Geoinfo, 112, 102869. doi: 10.1016/j.jag.2022.102869

TYPES OF EXPLAINABLE AI IN EARTH OBSERVATION



Gevaert (2022) Explainable AI for earth observation: A review including societal and regulatory perspectives, *Int. Journal of Applied EO and Geoinfo*, 112, 102869. doi: 10.1016/j.jag.2022.102869

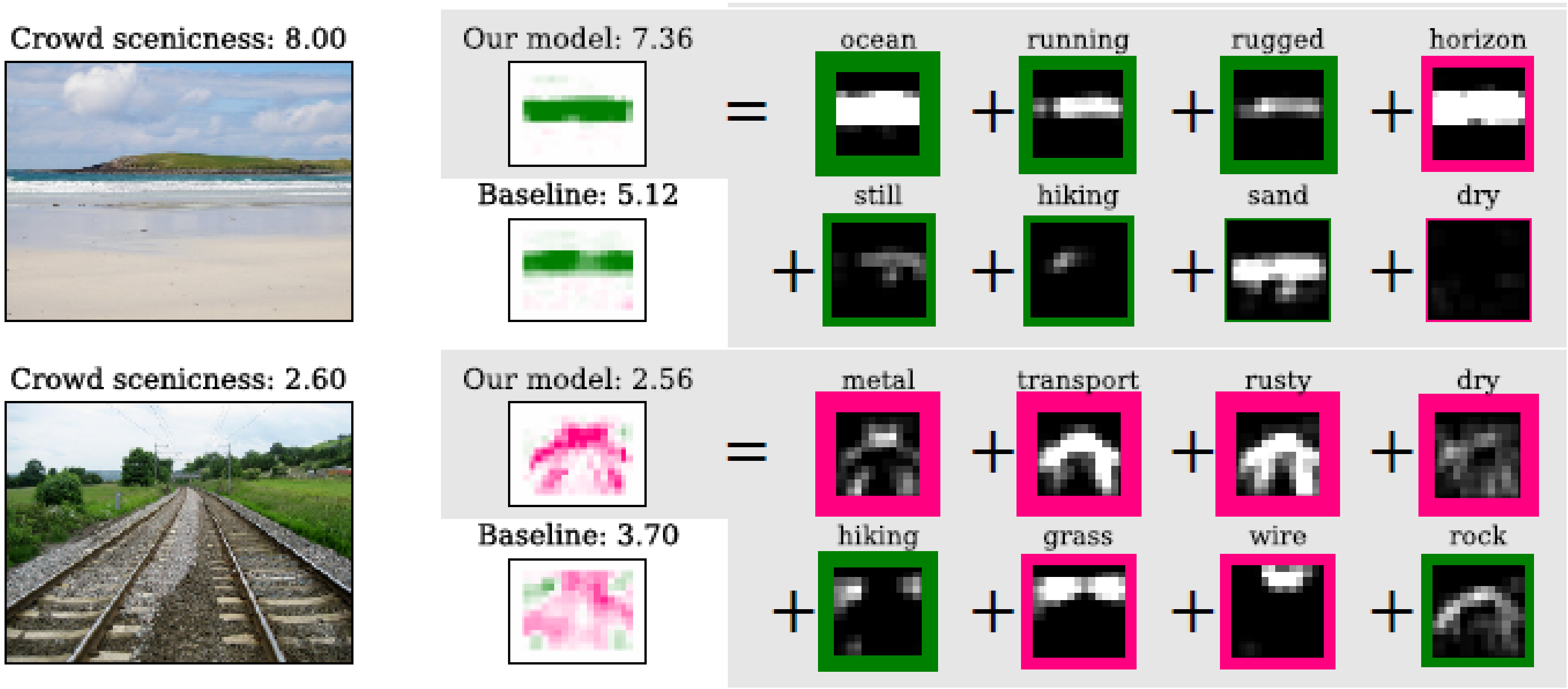
INTERPRETABLE MODELS



Marcos, D., Lobry, S., Tuia, D. 2019. Semantically Interpretable Activation Maps: what-where-how explanations with CNNs. ICCV 2019

UNIVERSITY OF TWENTE. ITC

INTERPRETABLE MODELS



Marcos, D., Lobry, S., Tuia, D. 2019. Semantically Interpretable Activation Maps: what-where-how explanations with CNNs. ICCV 2019

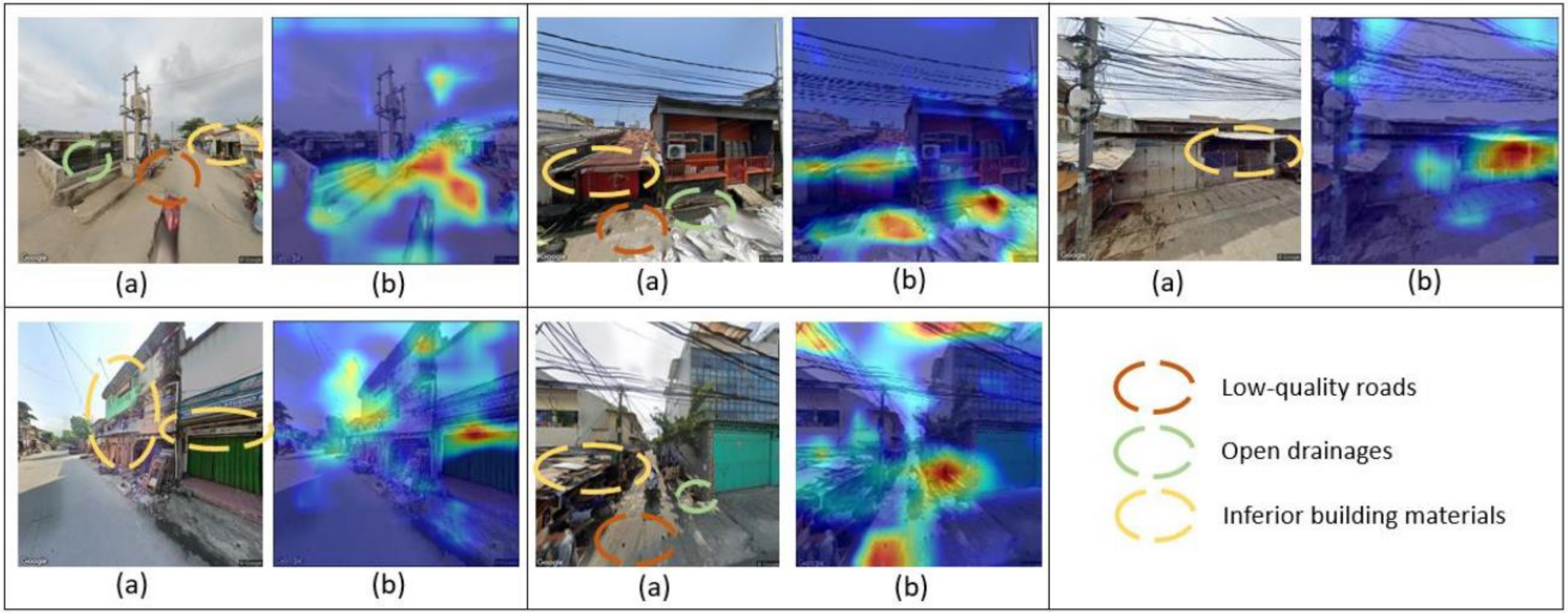
FEATURE SELECTION & IMPORTANCE

Features:

Color, texture, time, other sensors, elevation….

- **49% of the publications**
- Identify influential factors & induce sparsity.

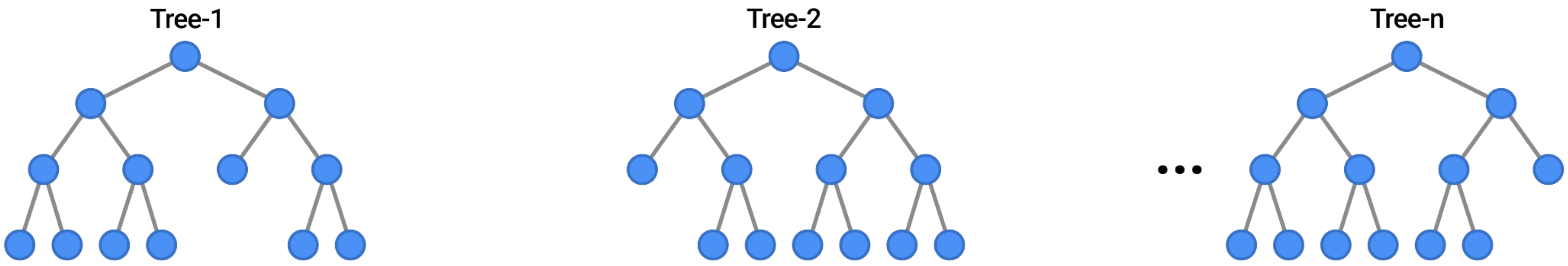
SALIENCY MAPS



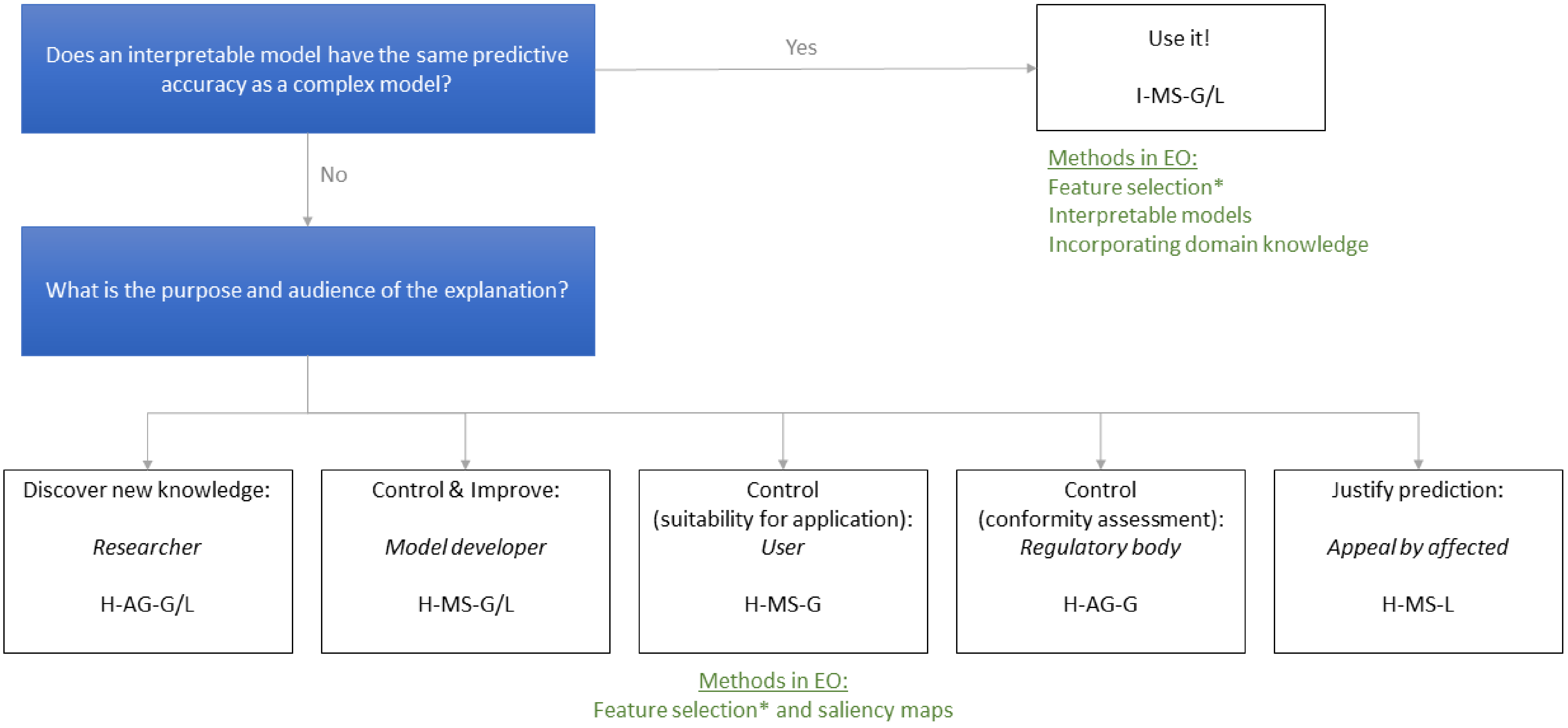
Source: Najmi, A. (2021) *Integrating Remote Sensing and Street View Images to Map Slums Using Deep Learning Approach* (MSc thesis).

CRITIQUES

ARE RANDOM FORESTS INTERPRETABLE?



UNIVERSITY OF TWENTE. | ITC



Gevaert (2022) Explainable AI for earth observation: A review including societal and regulatory perspectives, *Int. Journal of Applied EO and Geoinfo*, 112, 102869. doi: 10.1016/j.jag.2022.102869

EXPLAINABILITY – IS THIS MODEL SUITABLE?

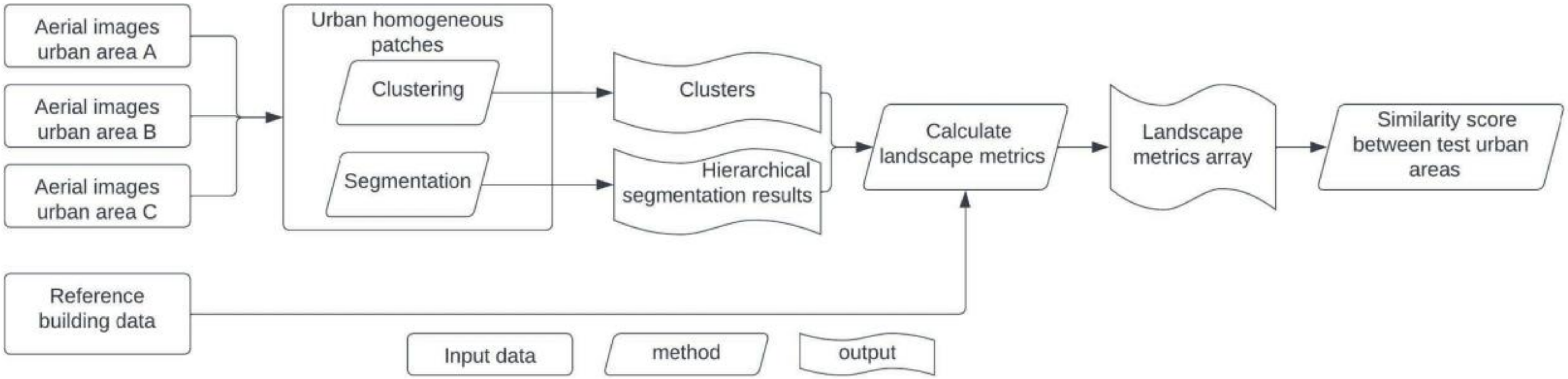
GENERALIZATION CAPABILITY

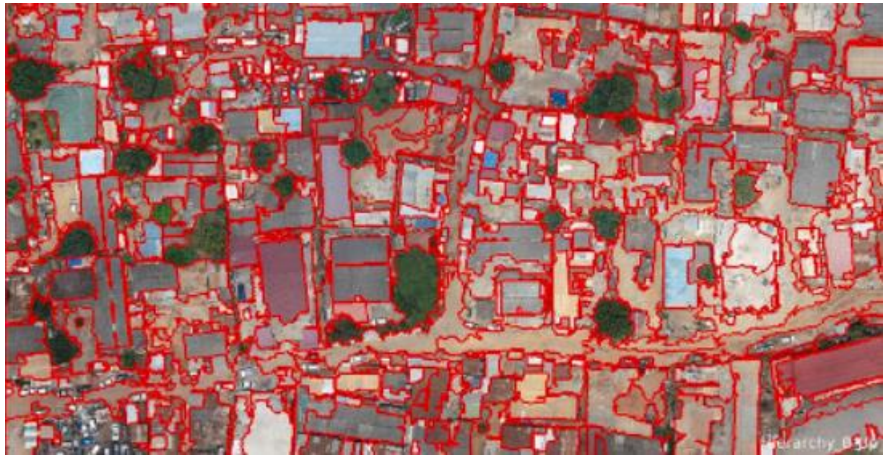
UNIVERSITY
OF TWENTE.

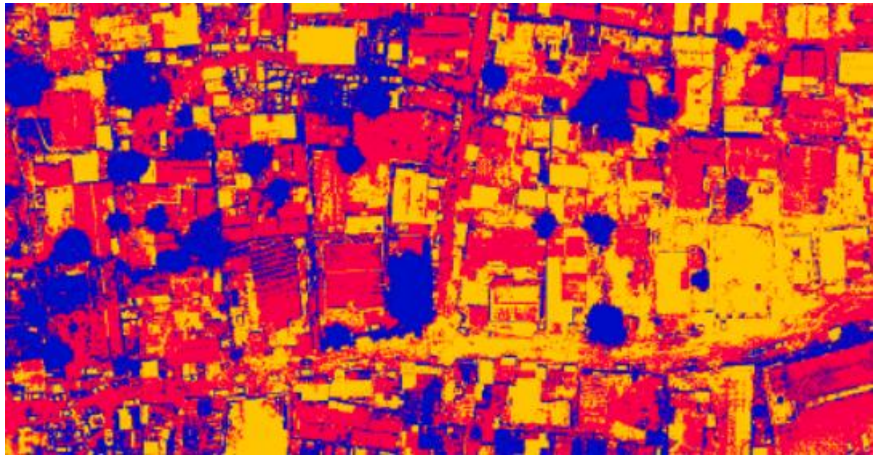
ACCRA | DAR ES SALAAM | ZANZIBAR

Gevaert, C.M. & Belgiu, M. (2022) Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *International Journal of Applied Earth Observation and Geoinformation,* 114.

UNIVERSITY OF TWENTE. ITC





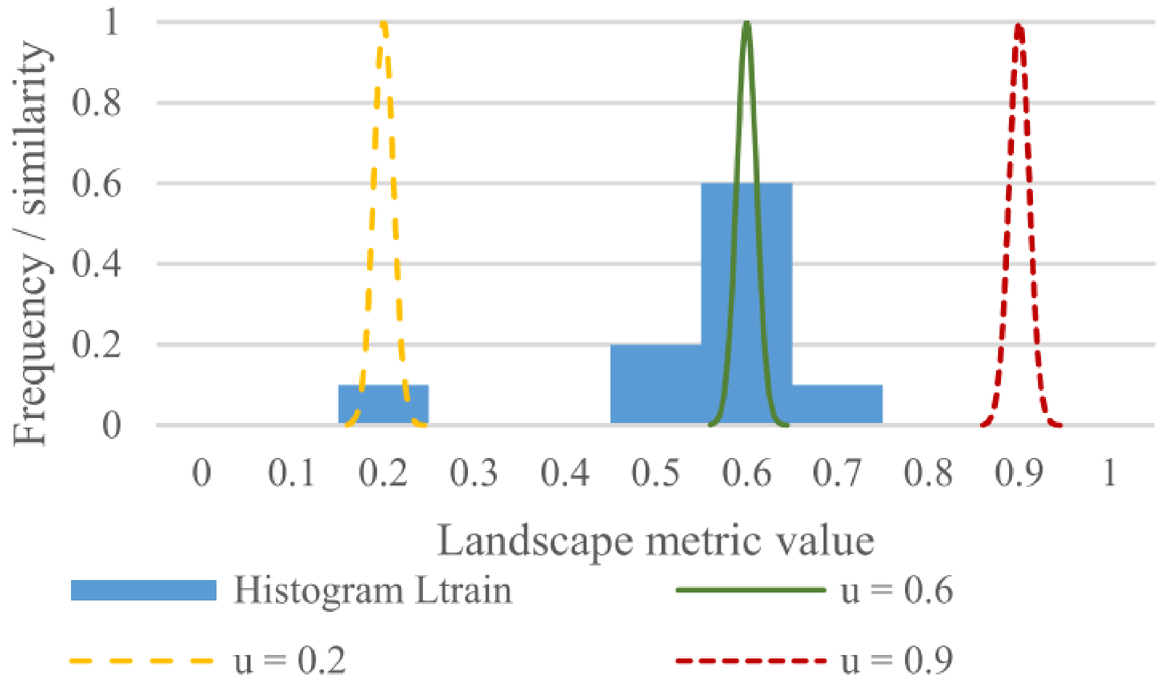


Gevaert, C.M. & Belgiu, M. (2022) Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *International Journal of Applied Earth Observation and Geoinformation,* 114.

UNIVERSITY OF TWENTE. | ITC

SIMILARITY SCORE

$$z_{uv} = \sum_{i=1}^{n} exp\left(\frac{-d^2(u, v_i)}{\sigma}\right)$$



Gevaert, C.M. & Belgiu, M. (2022) Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *International Journal of Applied Earth Observation and Geoinformation,* 114.

LANDSCAPE METRICS

Code	Name
NP	Number of patches
PD	Patch Density
ED	Edge Density
LSI	Landscape Shape Index
AREA_MN	Patch Area (Mean)
SHAPE_MN	Shape Index (Mean)
SHAPE_SD	Shape Index (Standard Deviation - SD)
FRAC_MN	Fractal Dimension Index (Mean)
FRAC_SD	Fractal Dimension Index (SD)
CONTIG_MN	Contiguity
CONTIG_SD	Contiguity Sd
CONTAG	Contagion
SHDI	Shannon's Diversity Index
SIDI	Simpson's Diversity Index
MSIDI	Modified Simpson's Diversity Index
SHEI	Shannon's Eveness Index

CORRELATION

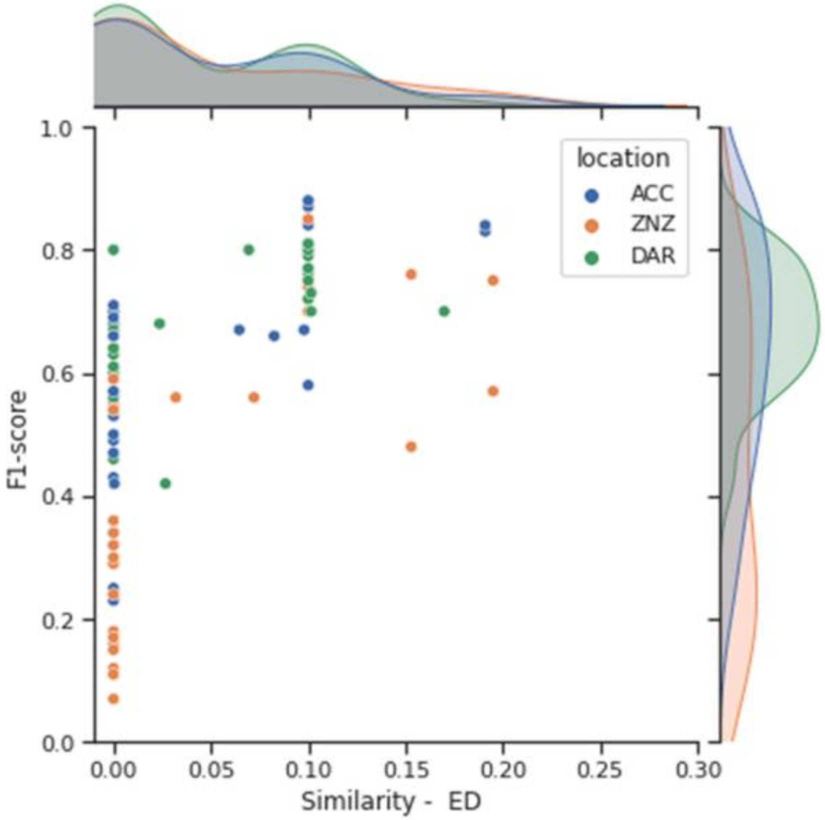
The correlation between the landscape metric similarity scores and the classification F1-scores, considering only Accra and Dar es Salaam. Cells highlighted in underline values indicate moderate correlations and bold values indicate strong correlations.

Landscape metric	Correlation to the F1-score					
	Segmentation			Clustering		Reference
	L1	L2	L3	k=2	k=3	
NP	0.66	0.56	0.58	0.66	0.59	**0.83**
PD	0.67	0.61	0.66	**0.71**	0.69	**0.81**
ED	0.58	0.57	**0.73**	**0.79**	**0.74**	**0.88**
LSI	0.50	0.30	0.46	**0.76**	0.66	**0.78**
AREA_AM	0.27	0.29	0.29	0.25	0.21	**0.82**
SHAPE_MN	0.67	0.54	0.09	0.10	0.40	0.66
SHAPE_SD	0.05	0.05	0.49	0.35	0.29	**0.72**
FRAC_MN	−0.08	0.32	0.2	**0.75**	**0.75**	0.59
FRAC_SD	−0.06	0.02	−0.03	**0.75**	**0.75**	0.12
CONTIG_MN	0.22	0.06	0.03	0.27	0.20	0.41
CONTIG_SD	0.27	0.26	−0.11	0.20	0.26	0.60
CONTAG	0.51	0.30	0.38	0.65	0.50	**0.81**
SHDI	–	–	–	0.14	0.28	**0.82**
SIDI	–	–	–	0.13	0.32	**0.82**
MSIDI	–	–	–	0.14	0.32	**0.82**
SHEI	–	–	–	0.14	0.28	**0.82**

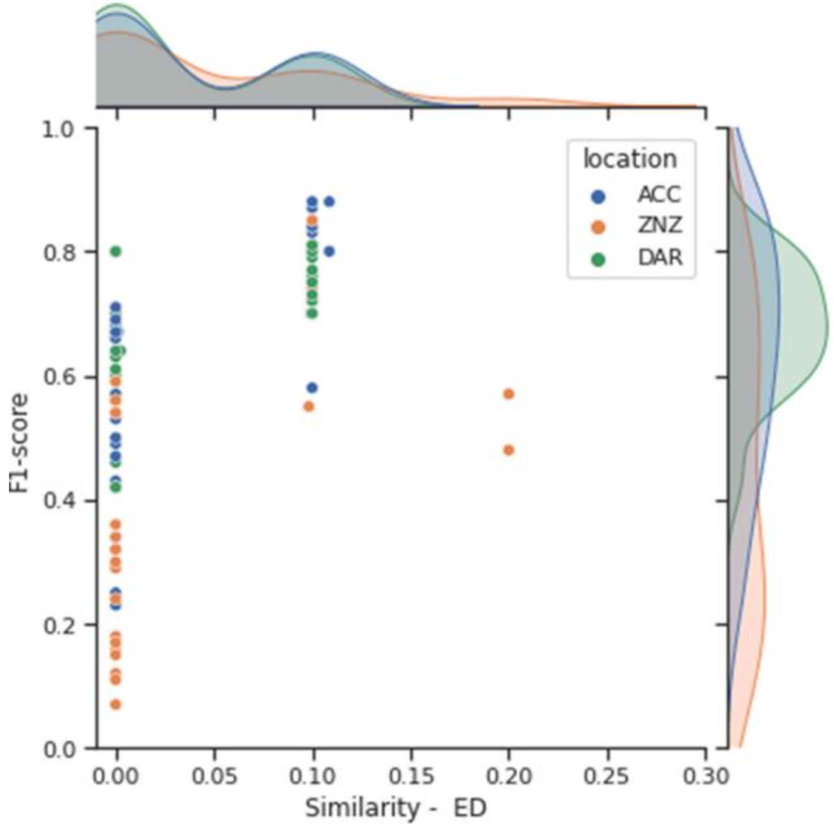
Gevaert, C.M. & Belgiu, M. (2022) Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *International Journal of Applied Earth Observation and Geoinformation*, 114.

UNIVERSITY OF TWENTE. ITC

Reference data



K-means clustering (k=2)



Gevaert, C.M. & Belgiu, M. (2022) Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics. *International Journal of Applied Earth Observation and Geoinformation,* 114.

UNIVERSITY OF TWENTE.

CONCLUSIONS

Biases are also present in EO data – methods for auditing are there, but the **challenge is identifying the sensitive attributes** and raising awareness that we need to audit for them.

Need **methods to predict generalizability**. – needed to provide explanations that legislation requires & understand when a model can be used.

Explanations in ML for EO not new, but changing

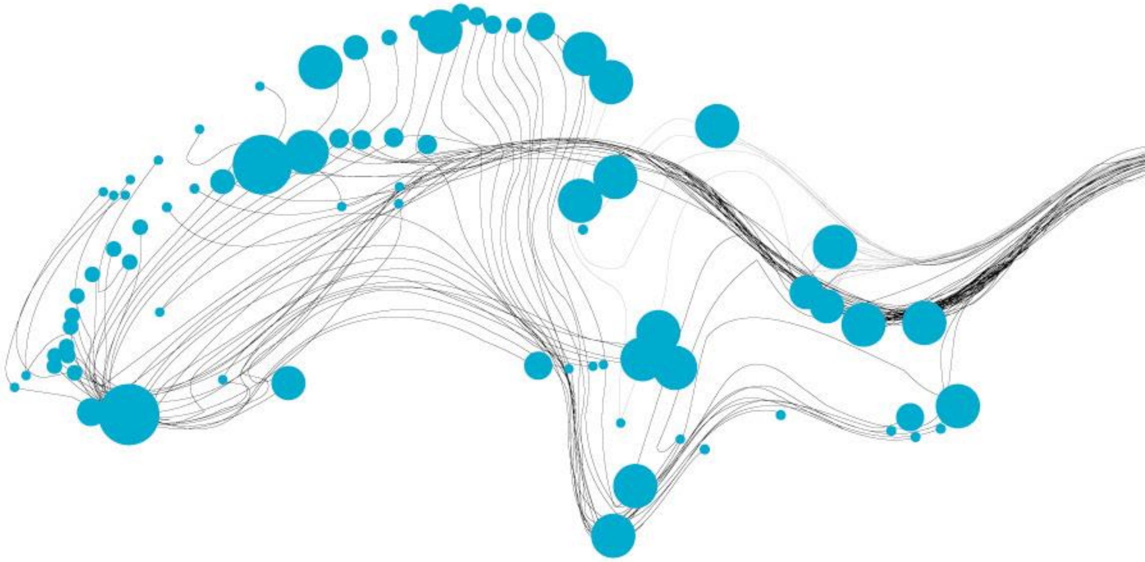
Limitations of explainability in EO:

- Which algorithms are considered **interpretable**
- Focus on **technical audience**
- **Lack of testing** of whether explanations are adequate

Kyoto University

October 4, 2023

RESPONSIBLE AI FOR EARTH OBSERVATION



CAROLINE GEVAERT

c.m.gevaert@utwente.nl

*This work is partially funded by the NWO-VENI project no. 18091

UNIVERSITY OF TWENTE. | ITC