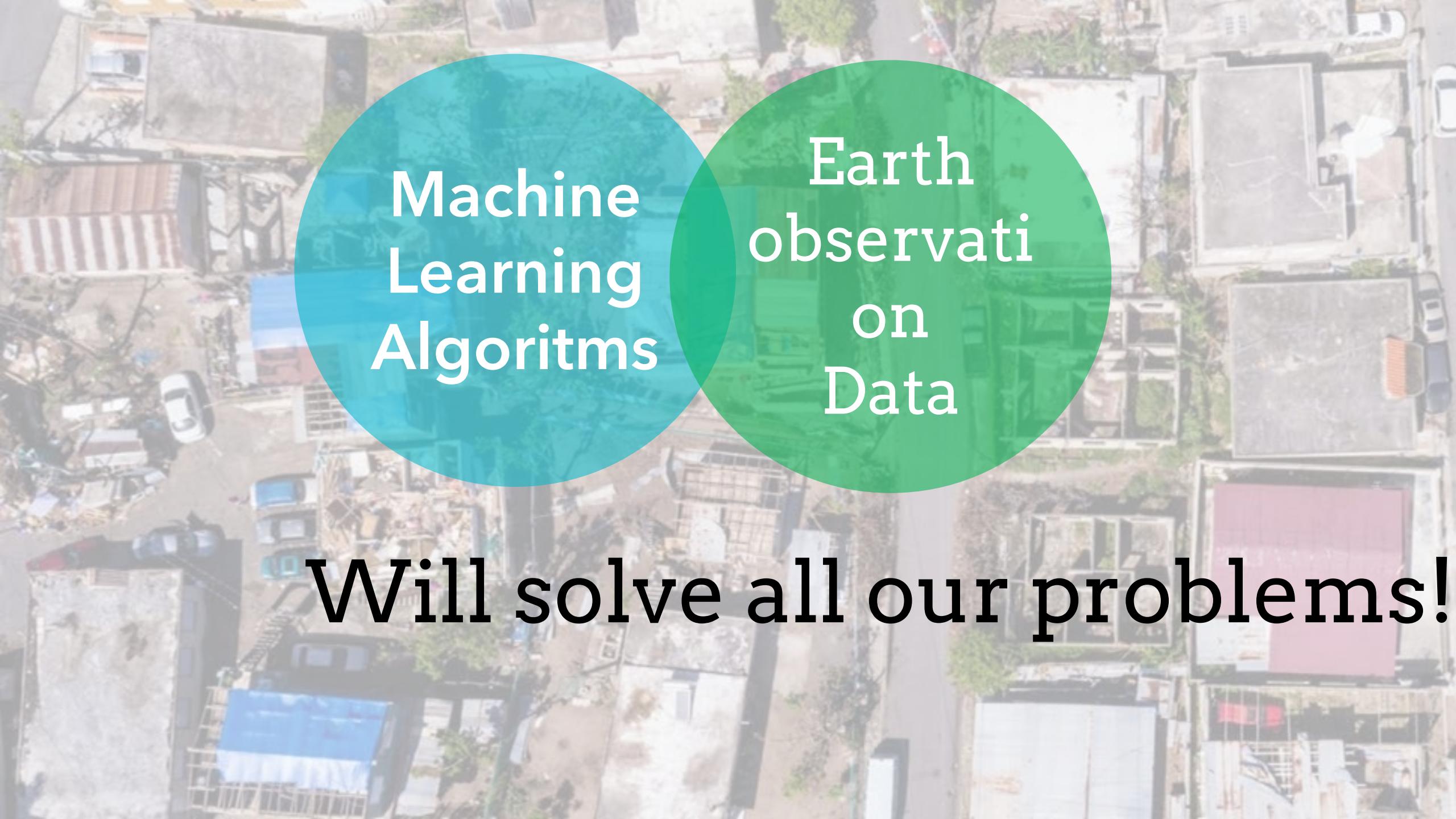


The (un)limited potential of machine learning for Earth Observation

Stef Lhermitte



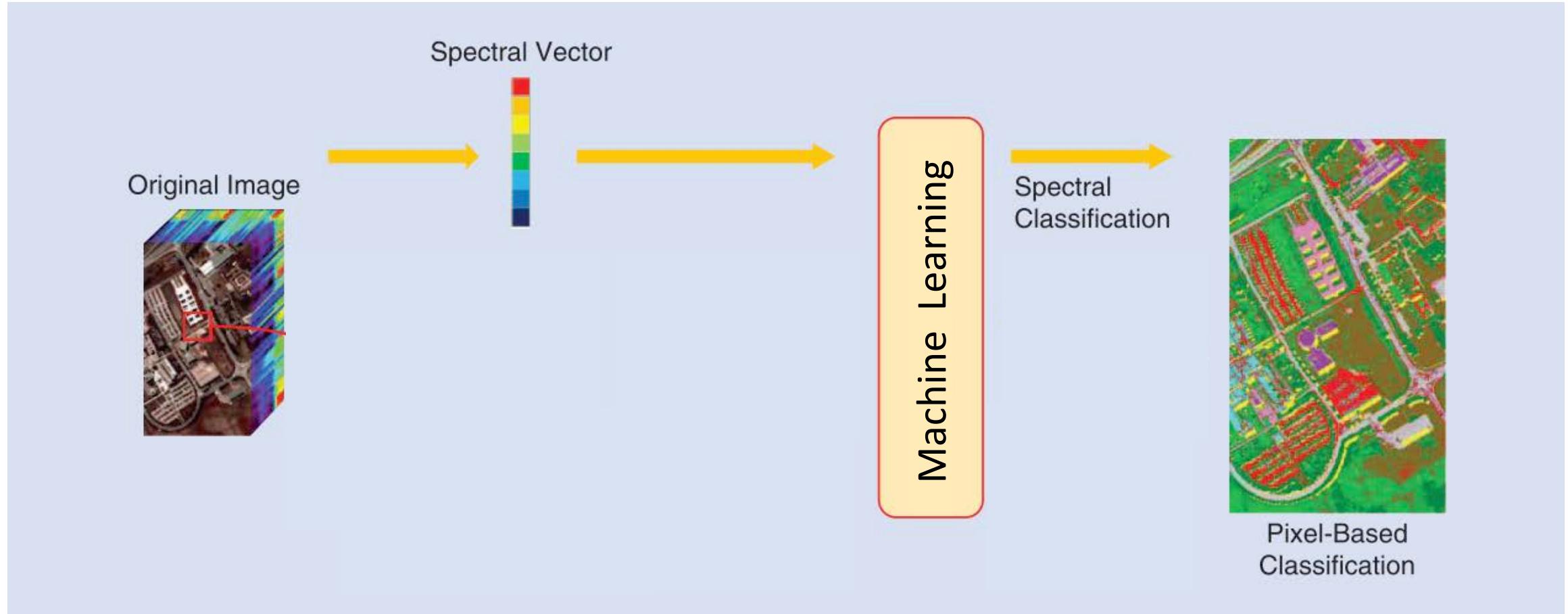
A circular graphic overlay on a background aerial photograph of a city. The left half of the circle is teal, and the right half is green. The text 'Machine Learning Algorithms' is in white on the teal side, and 'Earth observation Data' is in white on the green side.

Machine
Learning
Algorithms

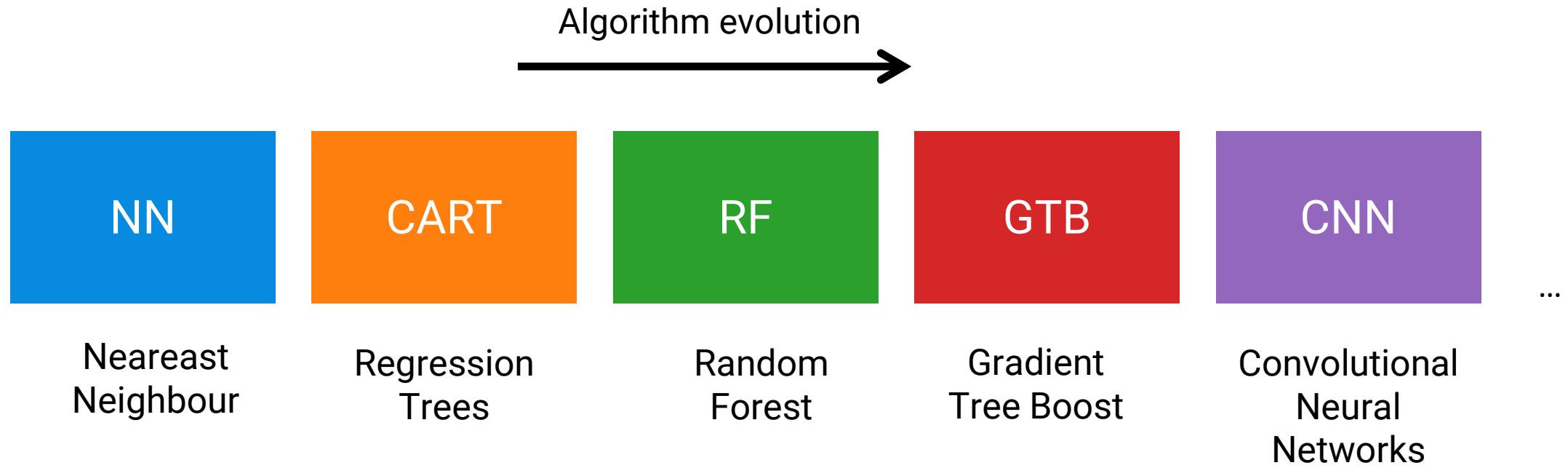
Earth
observati
on
Data

Will solve all our problems!

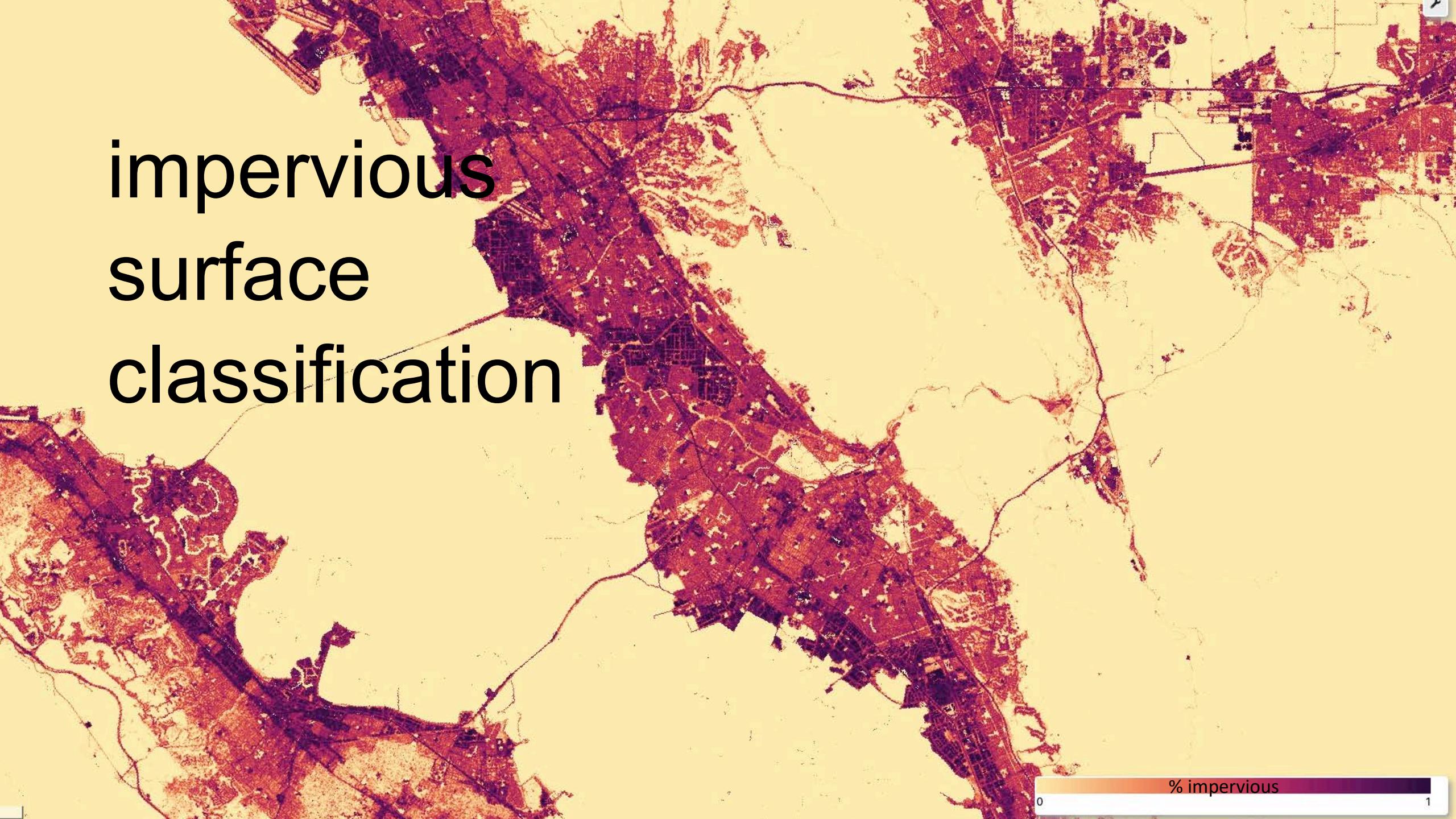
Hooray for the unlimited potential of ML methods



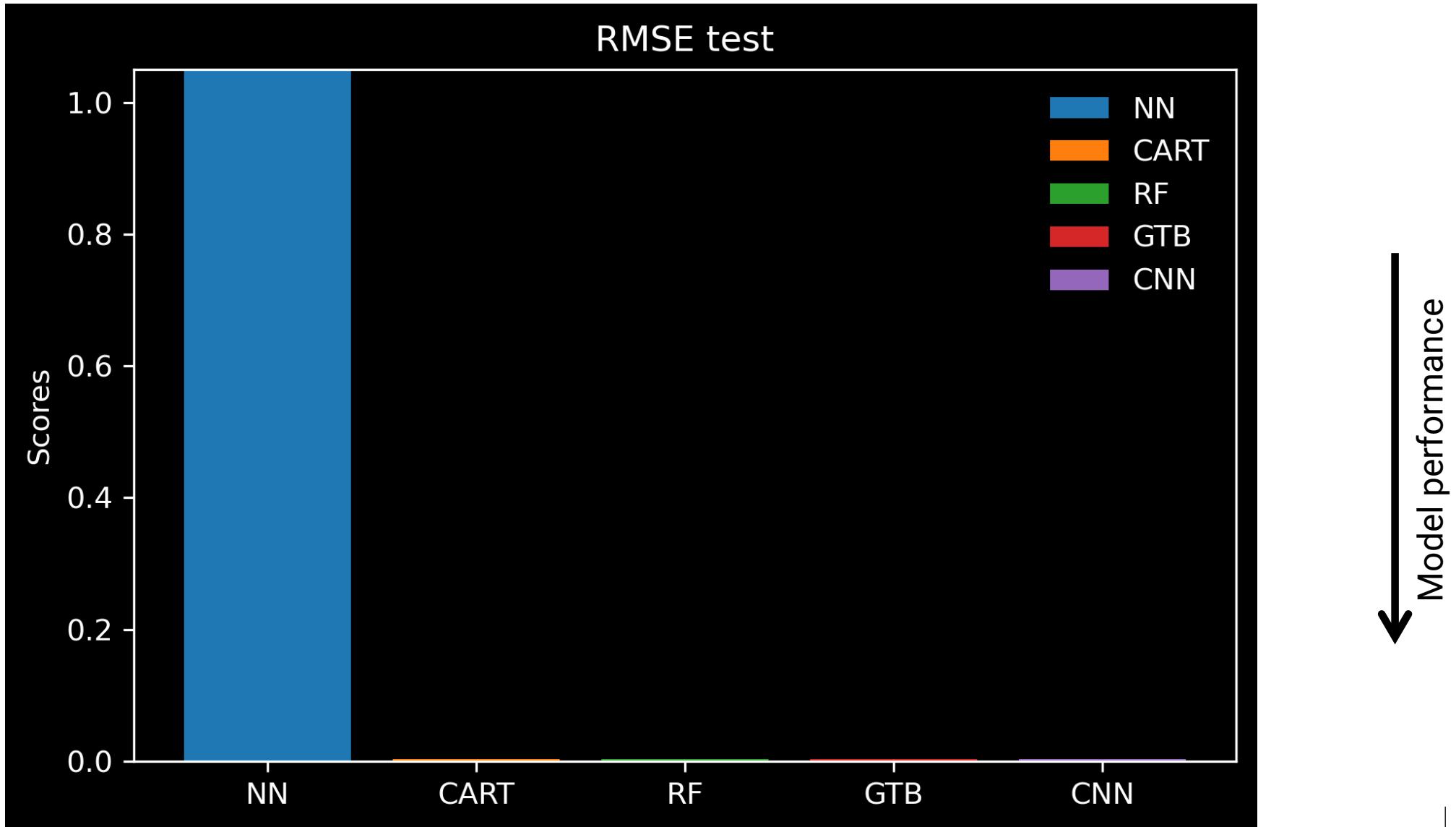
ML method evolution



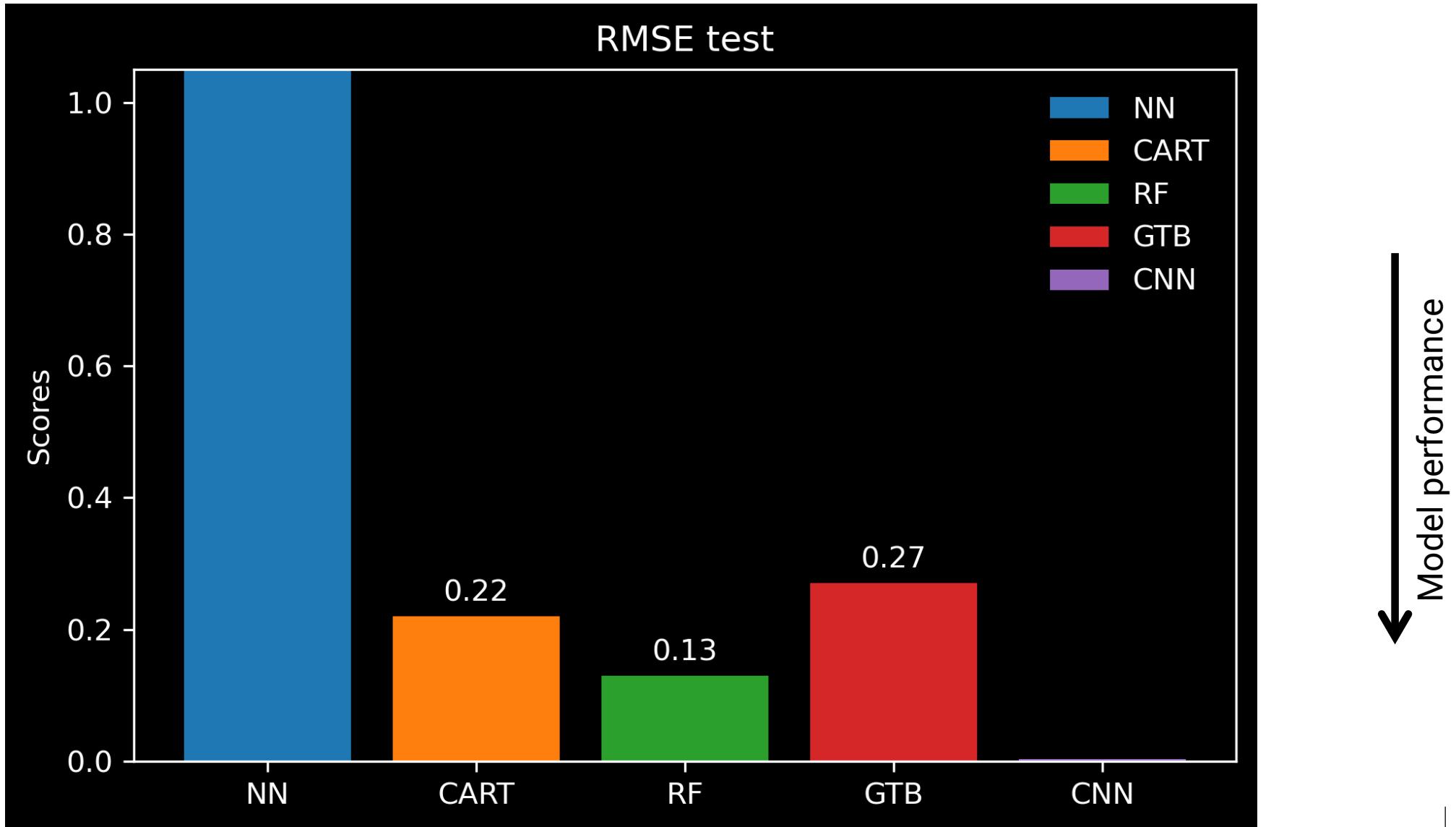
impervious surface classification



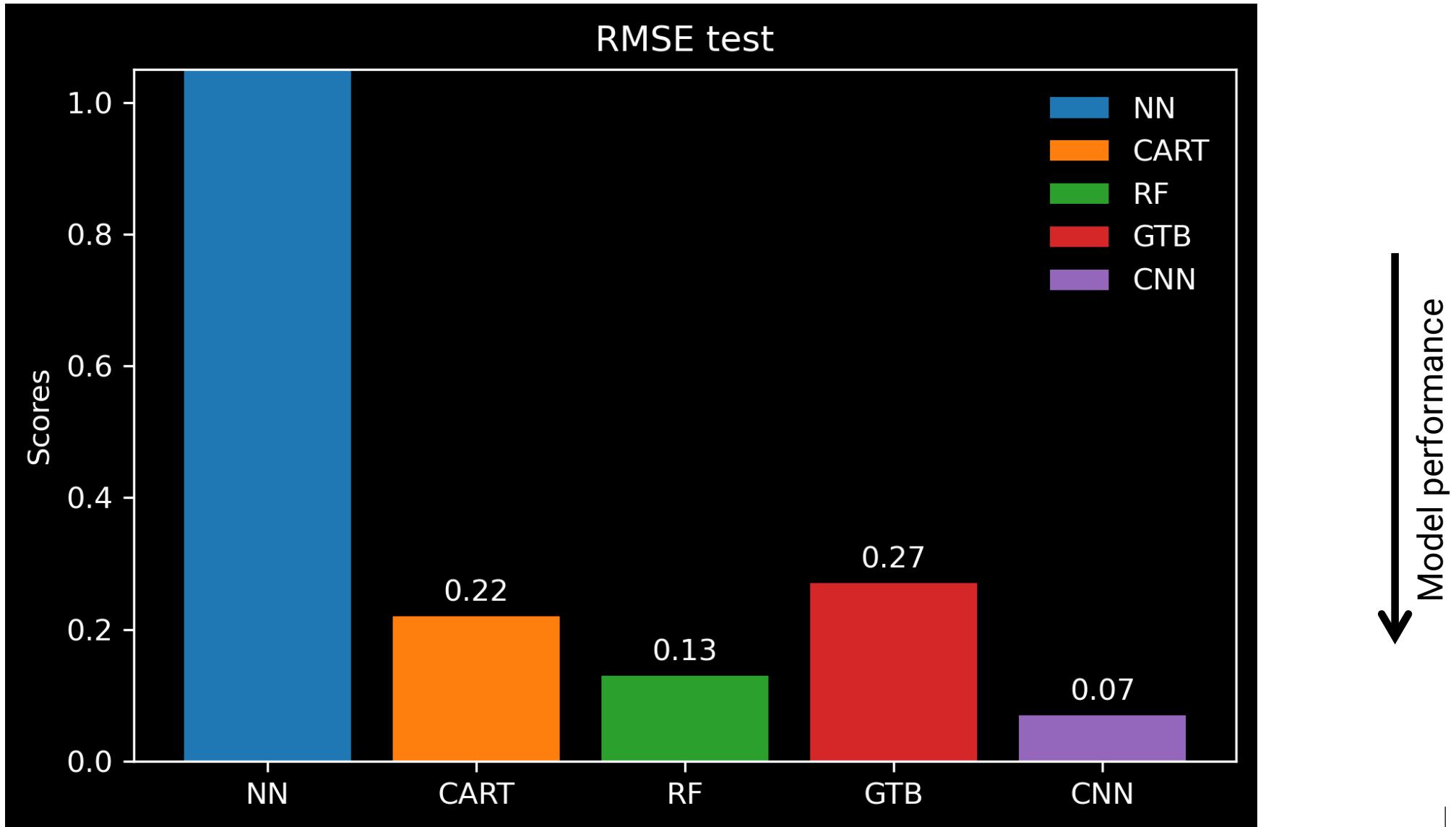
How good are the ML algorithms?



How good are the ML algorithms?



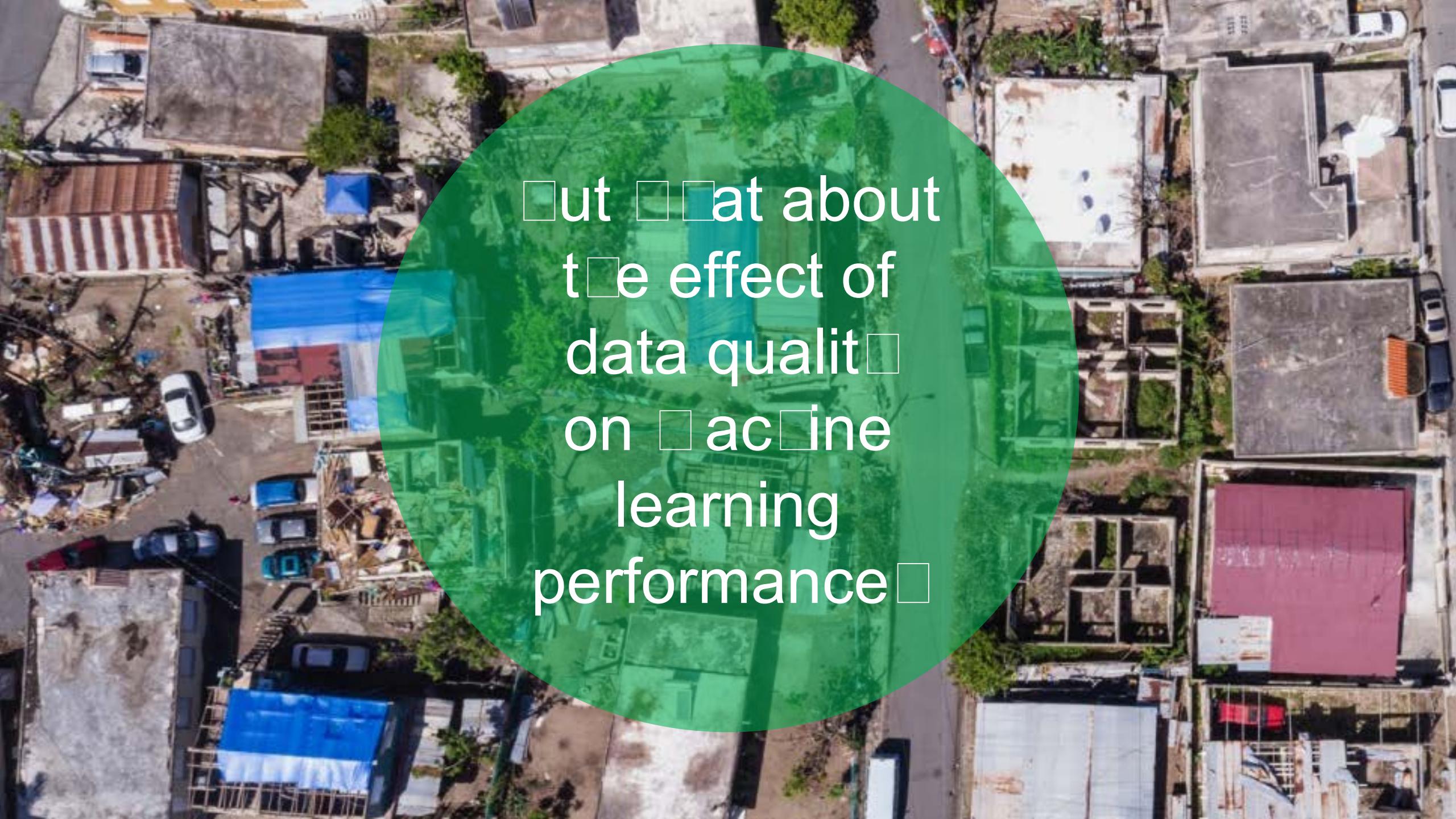
How good are the ML algorithms?





A large, semi-transparent teal circle is positioned in the center of the slide, containing the following text:

- ☐ et☐ods are
constantl☐
improving
- ☐
- ☐nlimited
potential



A large green circular overlay covers the center of the image, containing white text. The text discusses the effect of data quality on machine learning performance.

Out what about
the effect of
data quality
on machine
learning
performance



A large green circular overlay covers the center of the image, containing white text. The text reads "Ne□ Data" on top and "ne□ problems" below it, with small square icons integrated into the letters.

Ne□ Data
ne□ problems

New data, new problems: e.g. shadows





Landsat-8



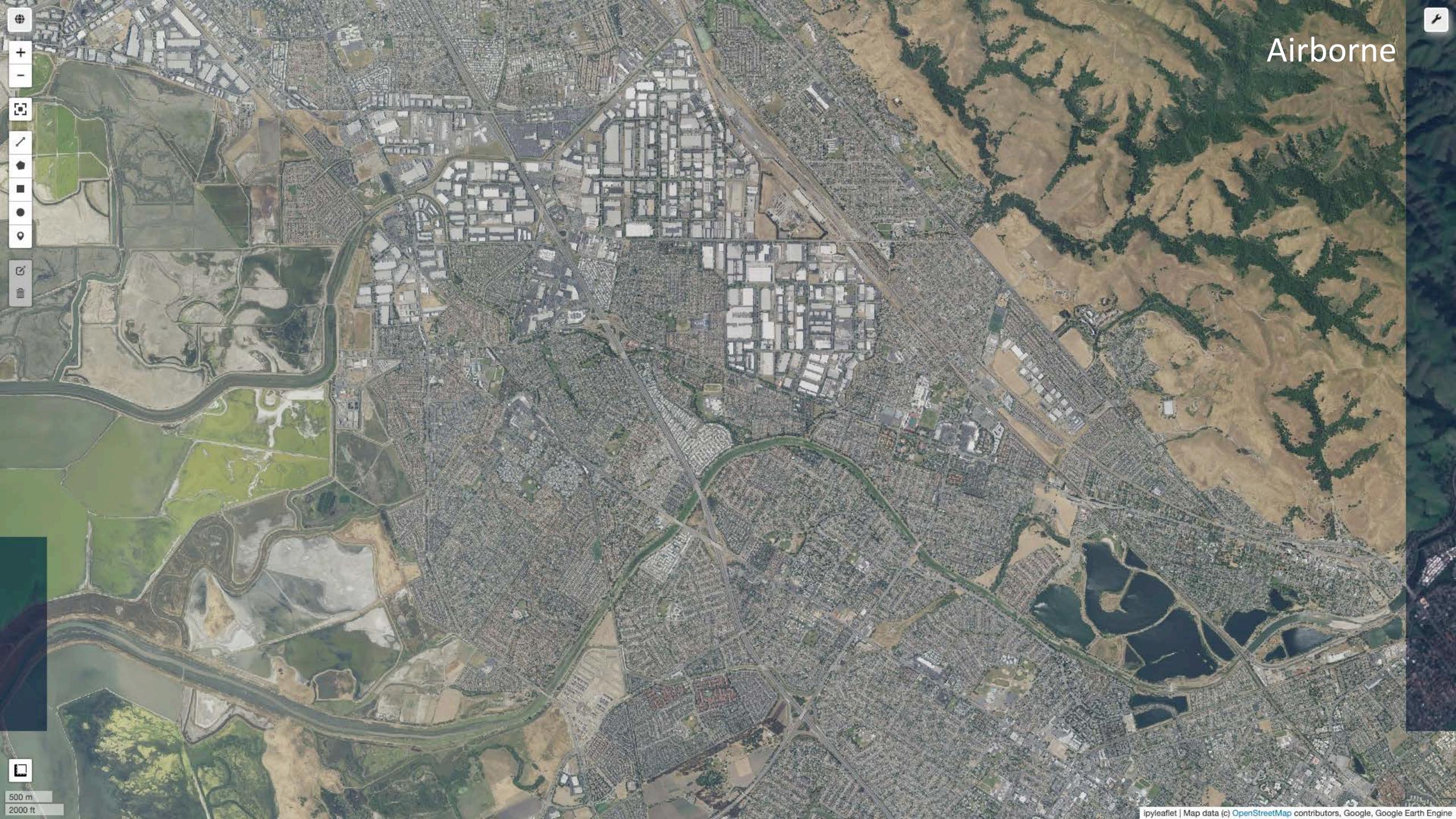
500 m
2000 ft



Sentinel-2

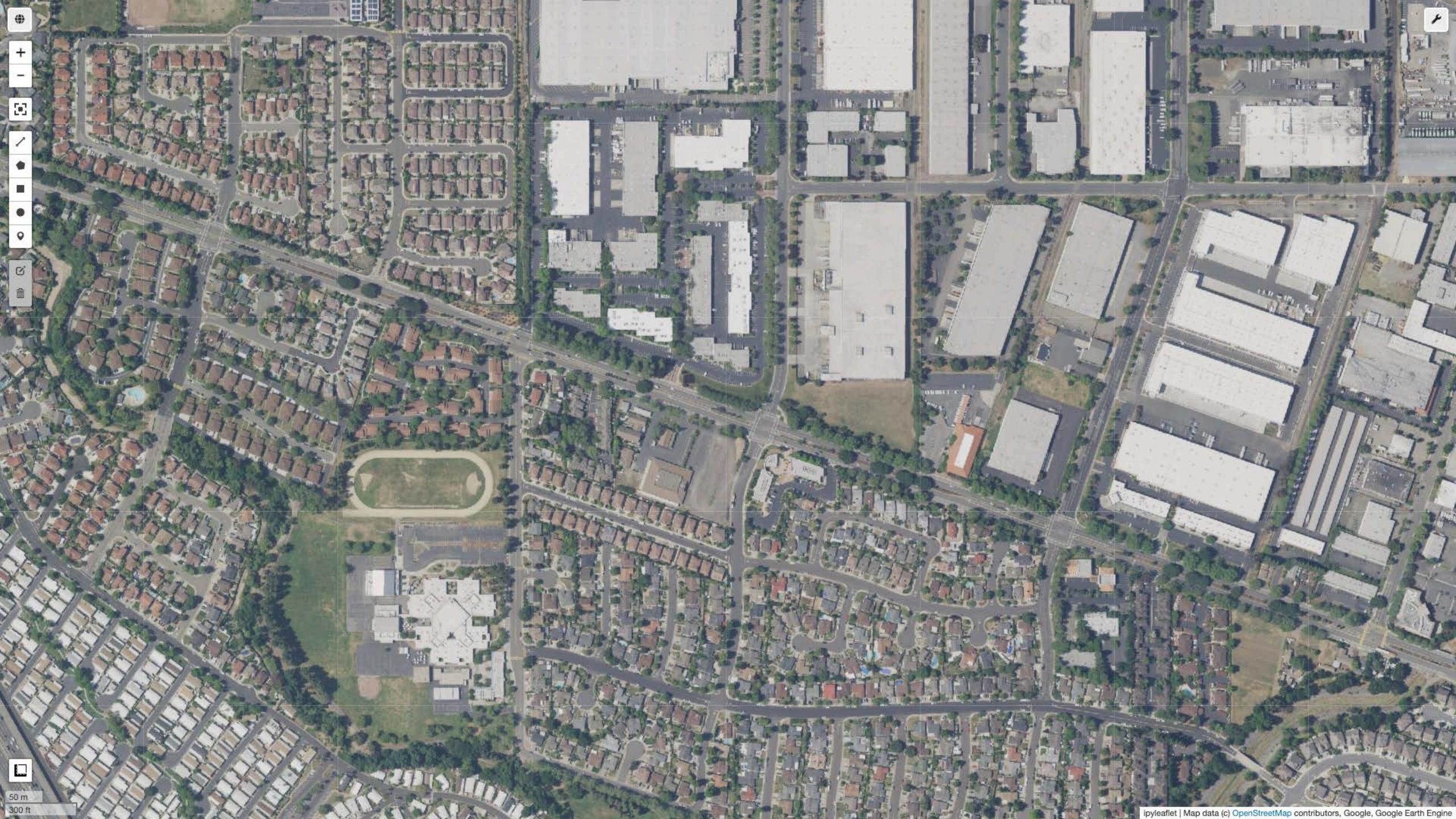


500 m
2000 ft

Airborne

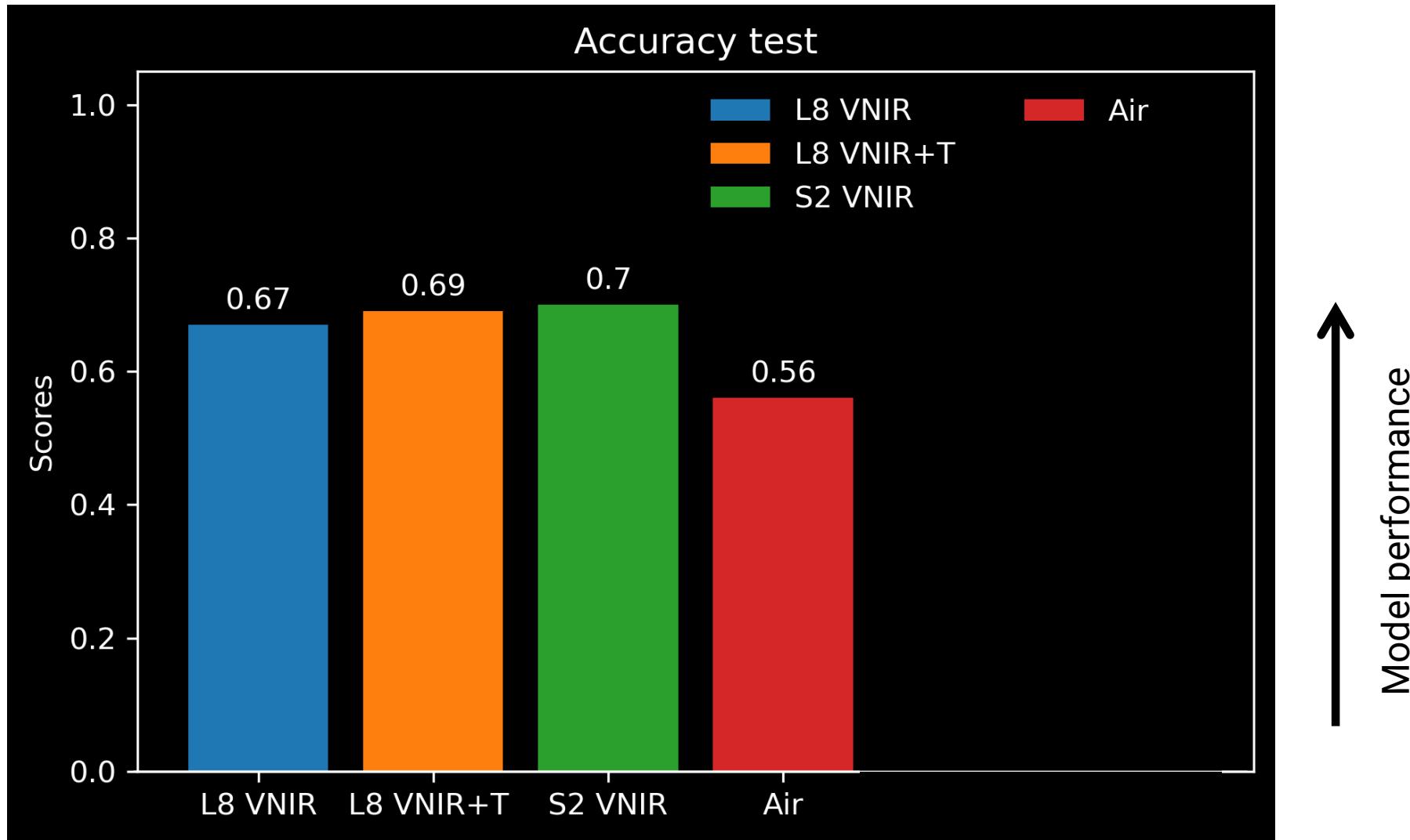
500 m

2000 ft

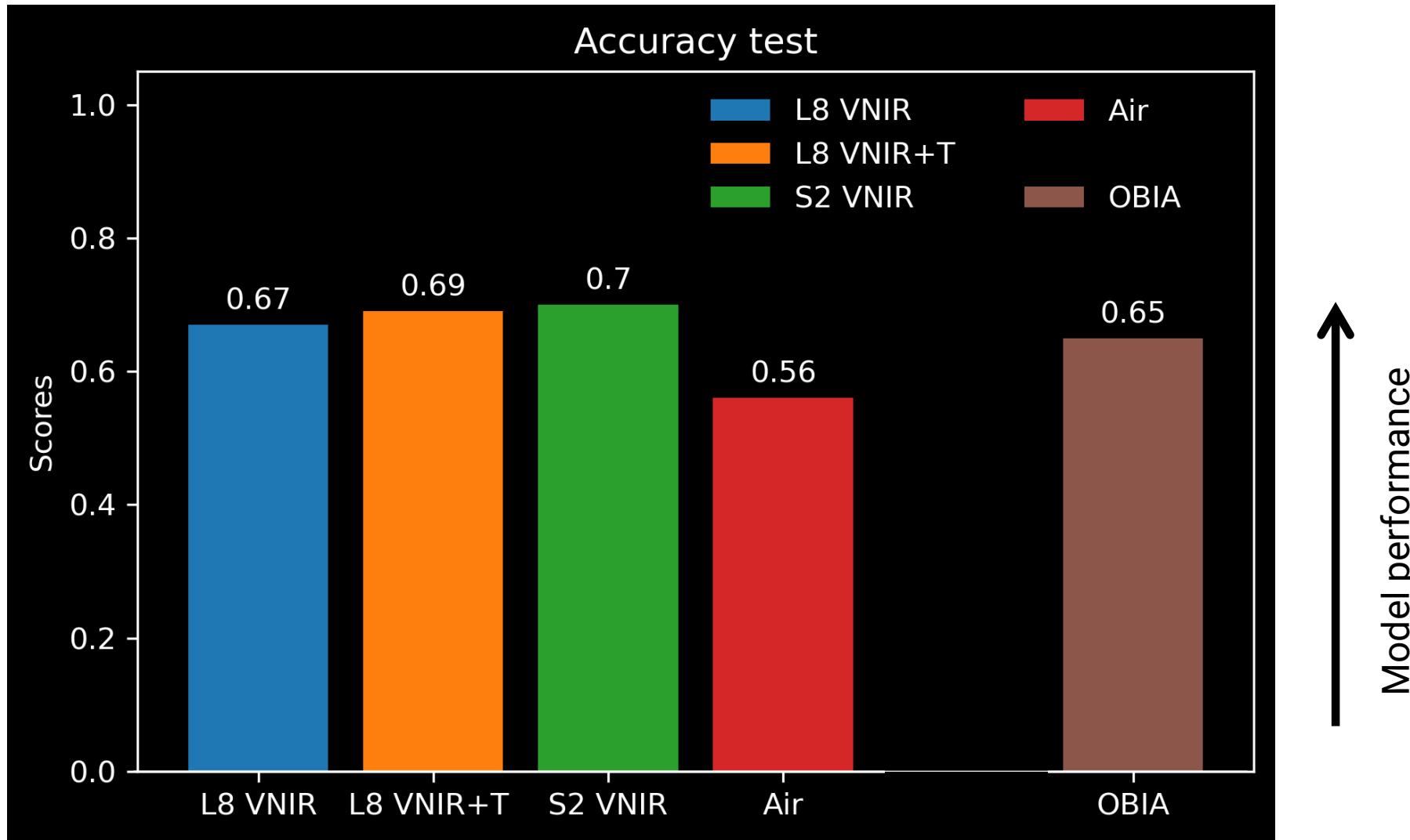


50 m
300 ft

New data problems: e.g. shadows



New data problems: e.g. shadows



New data problems: change detection for disaster management

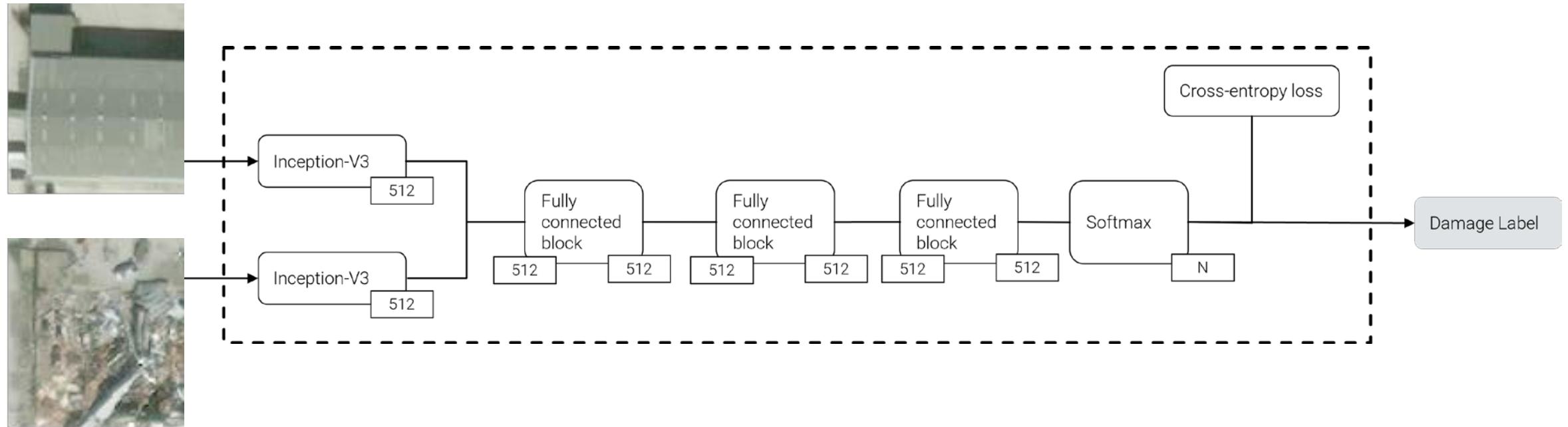
Pre-event



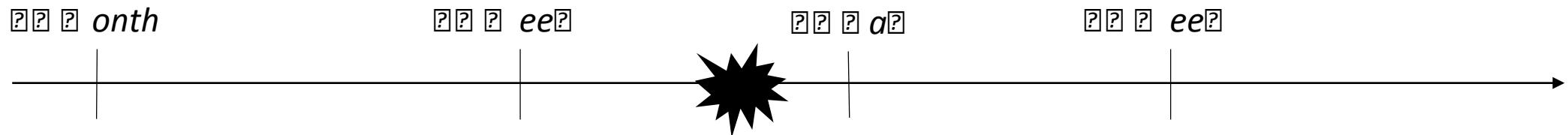
Post-event



The 510 Caladrius model



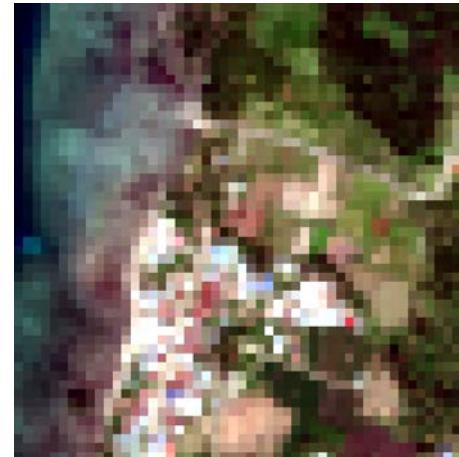
New data problems: time urgency



Highres



Lo? Pres



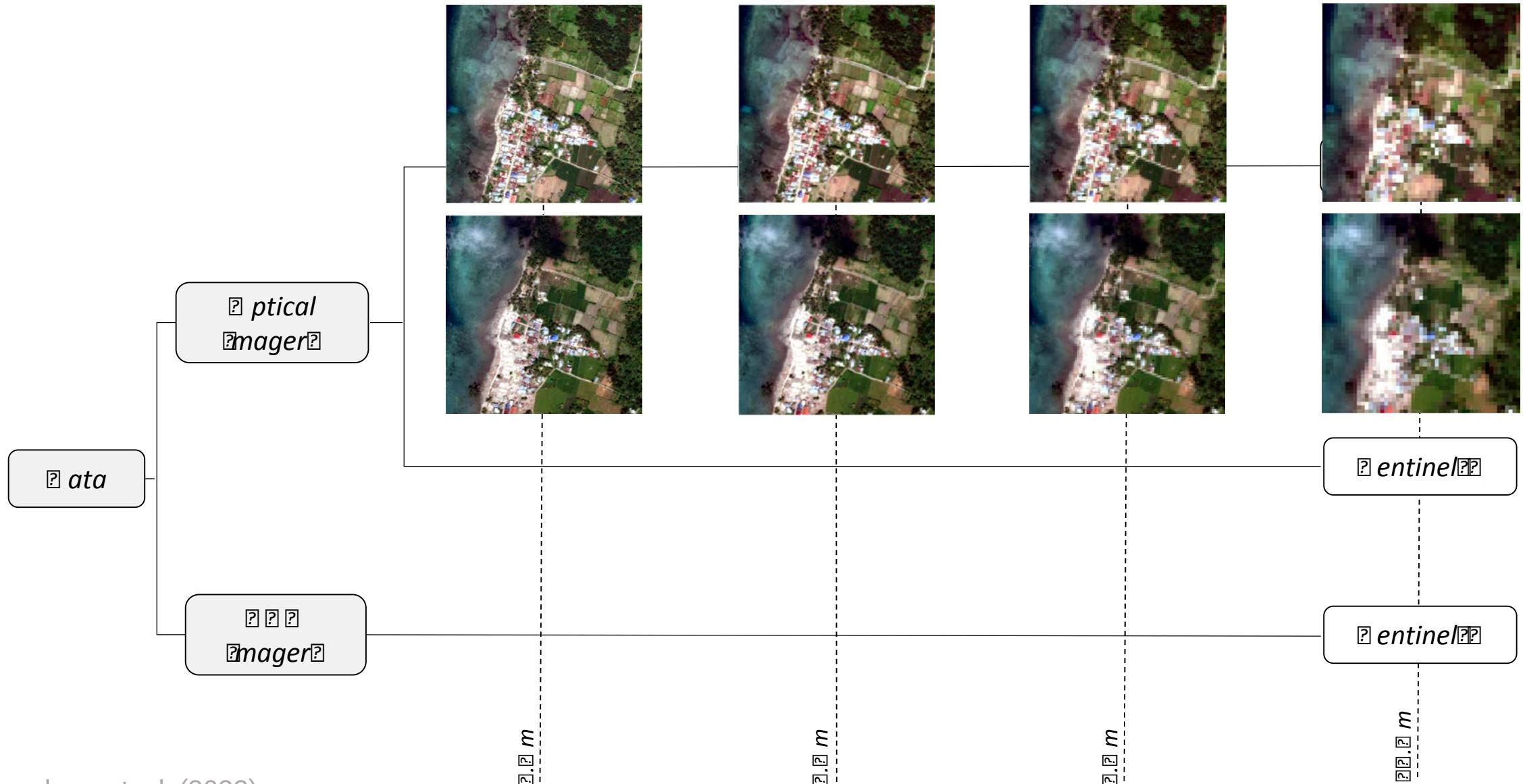
Lo? Pres



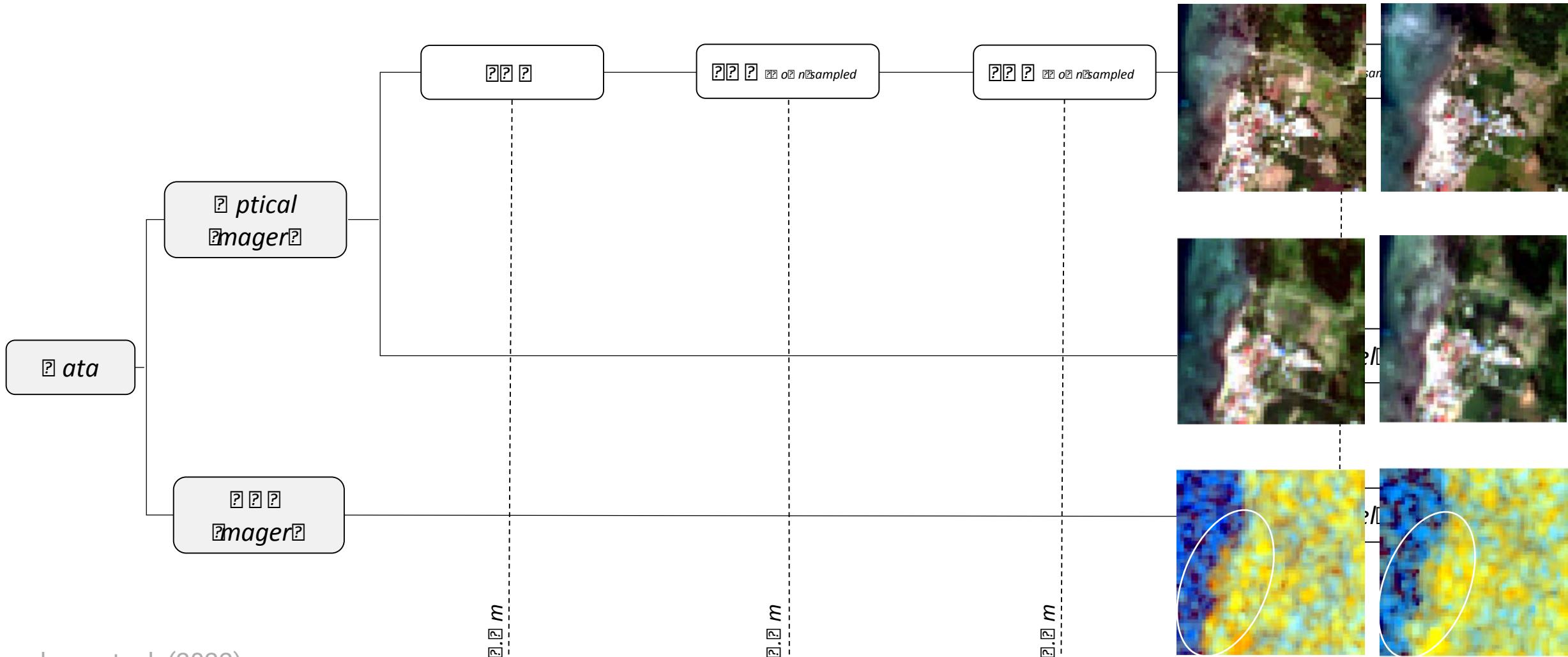
Highres



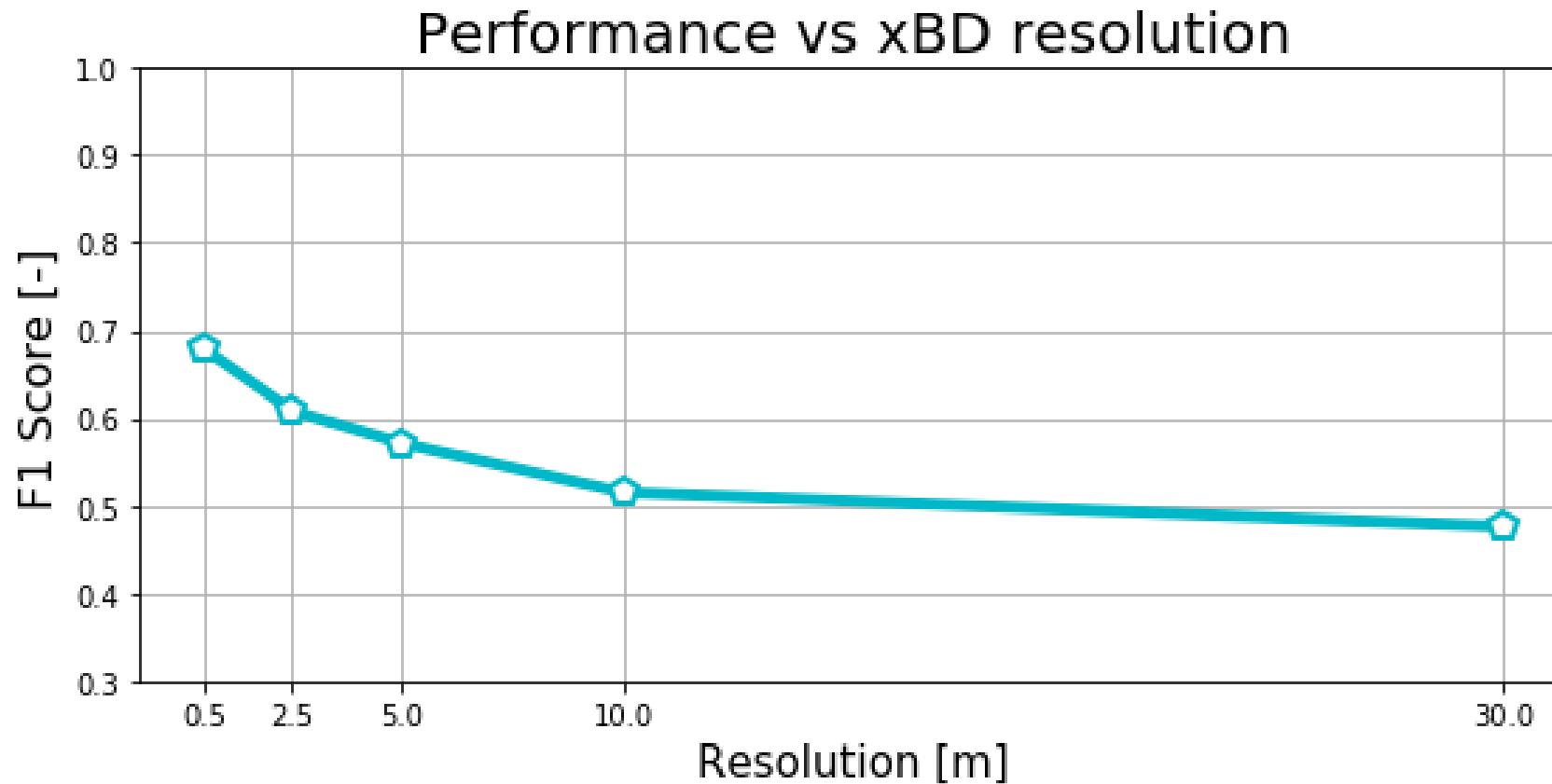
The input data is originated from the xBD, Sentinel-2 and Sentinel-1 dataset



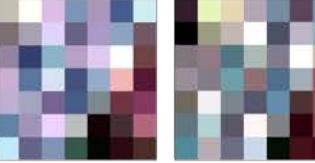
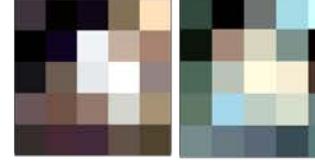
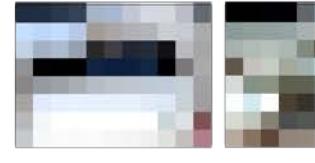
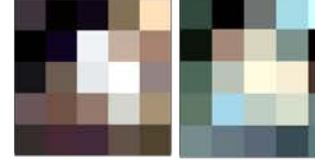
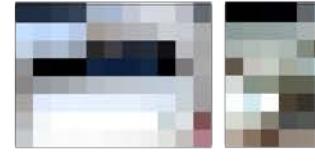
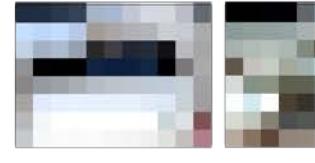
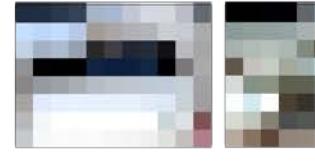
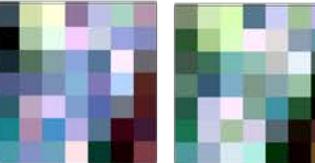
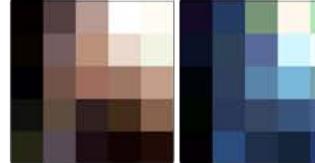
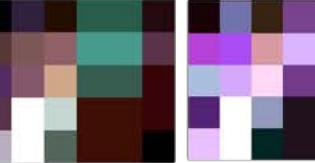
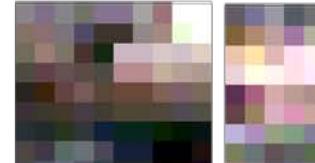
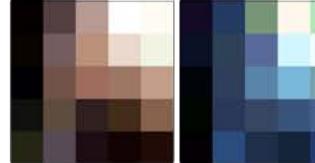
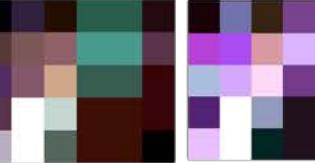
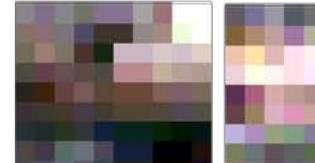
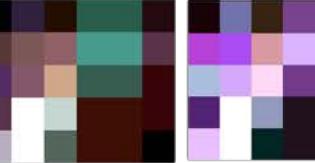
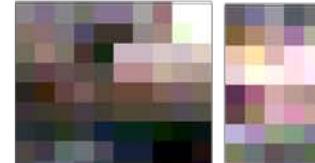
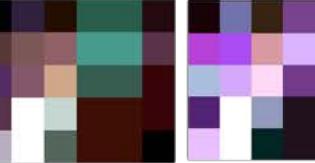
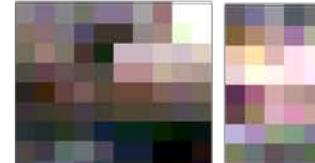
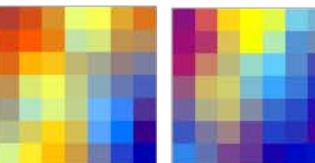
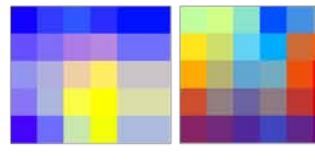
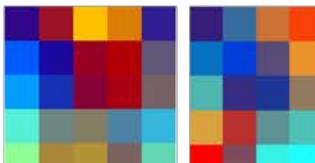
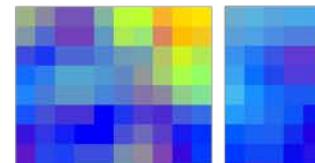
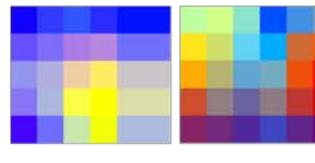
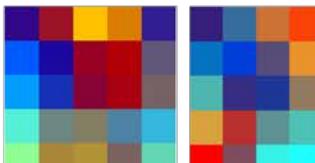
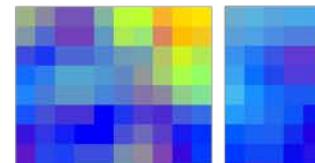
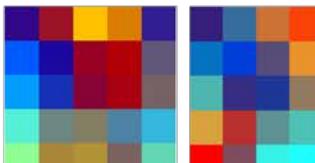
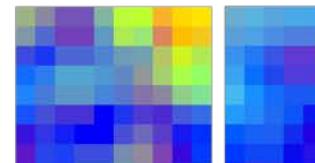
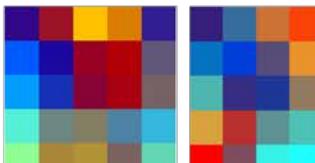
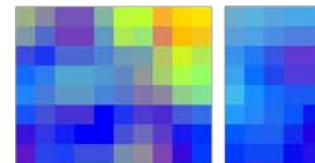
The input data is originated from the xBD, Sentinel-2 and Sentinel-1 dataset



Data resolution matters ...



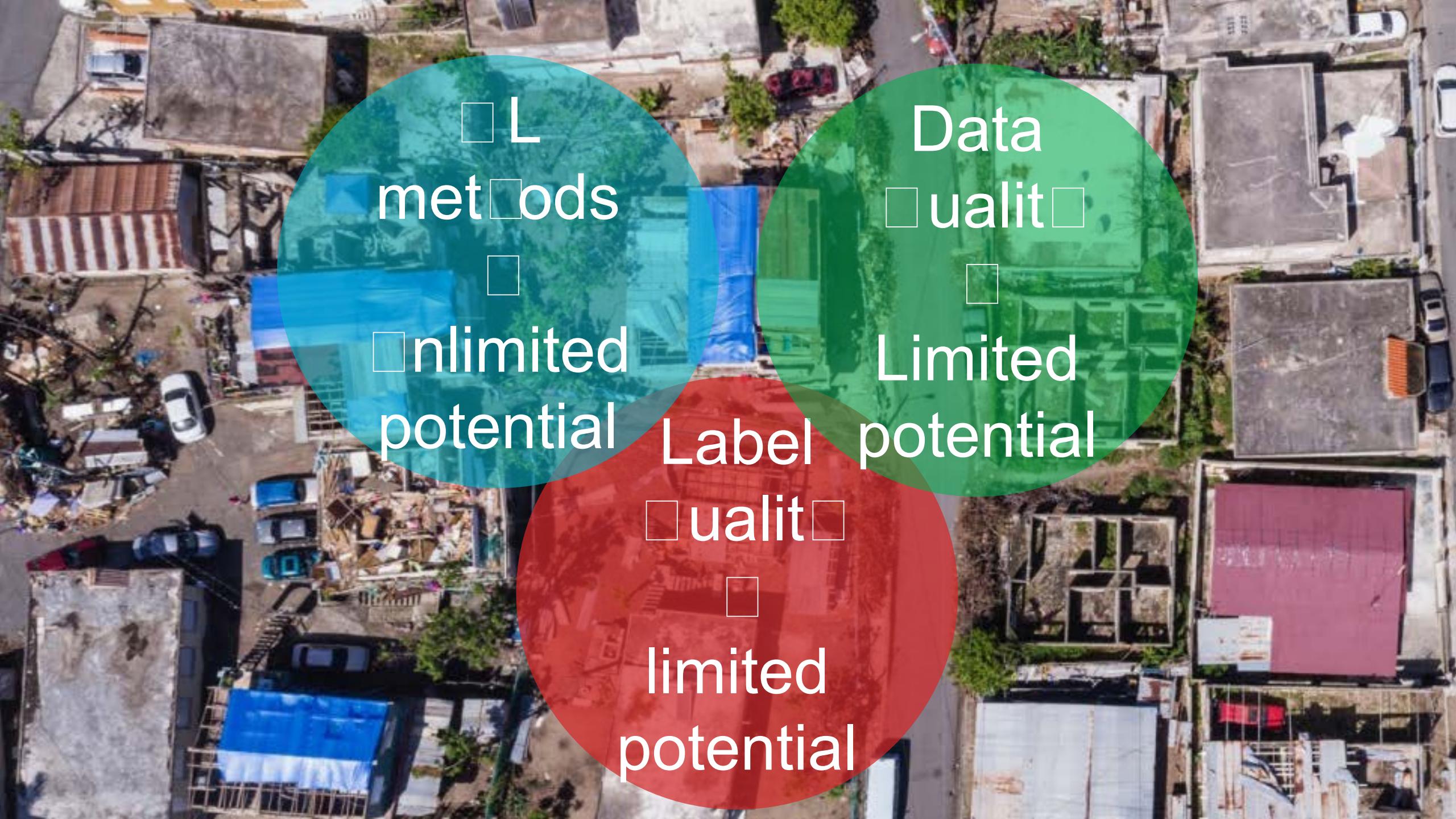
Data type matters

Example	Example 1	Example 2	Example 3	Example 4	
Label	No-damage	Minor-damage	Major-damage	Destroyed	
xBD 0.5m Optical data	 ✓  ✓  ✓  ✓  ✓	 ✓  ✓  ✓  ✓	 ✓  ✓  ✓	 ✓  ✓  ✓	0.68 acro ??
xBD 10.0m Optical data	 ✗  ✗  ✓  ✗  ✗	 ✓  ✗  ✗  ✗	 ✗  ✗  ✗	 ✗  ✗  ✗	0.52
Sentinel-2 Optical data	 ✓  ✓  ✗  ✗  ✗	 ✗  ✗  ✗  ✗	 ✗  ✗  ✗	 ✗  ✗  ✗	0.44
Sentinel-1 SAR data	 ✗  ✗  ✓  ✗  ✗	 ✓  ✗  ✗  ✗	 ✗  ✗  ✗	 ✗  ✗  ✗	0.32

✓ Correct ✗ Incorrect

... but not always that much

		Post-Event			
		0.5m	2.5m	5.0m	10.0m
Pre-Event	xBD 0.5m	0.65	0.63	0.61	0.57
	xBD 2.5m	X	0.61	0.58	0.57
	xBD 5.0m	X	X	0.57	0.53
	xBD 10.0m	X	X	X	0.52



□ L
met□ods

□nlimited
potential

Data
□ualit□

Limited
potential

Label

□ualit□

□
limited
potential

Conclusion

ML methods offer promising methodological opportunities but are no magical solutions.

ML methods require application/data/label-specific development

Even the best ML methods suffer from “*garbage in = garbage out*”