



Manaaki Whenua
Landcare Research

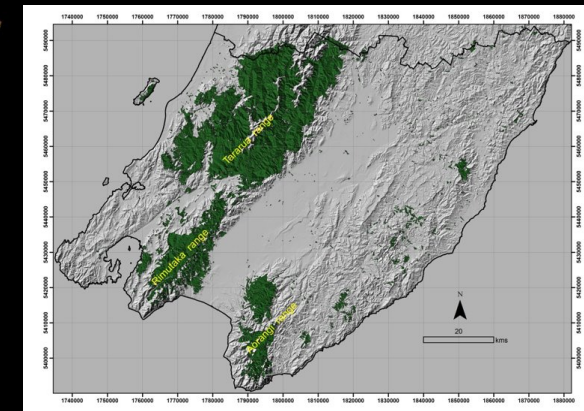
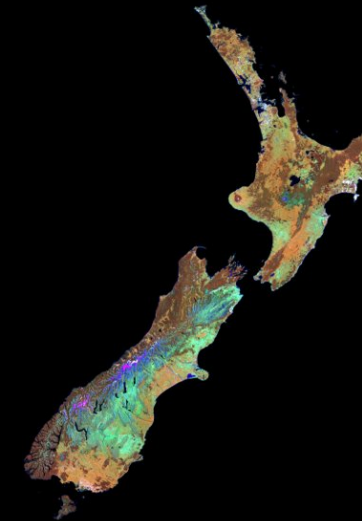
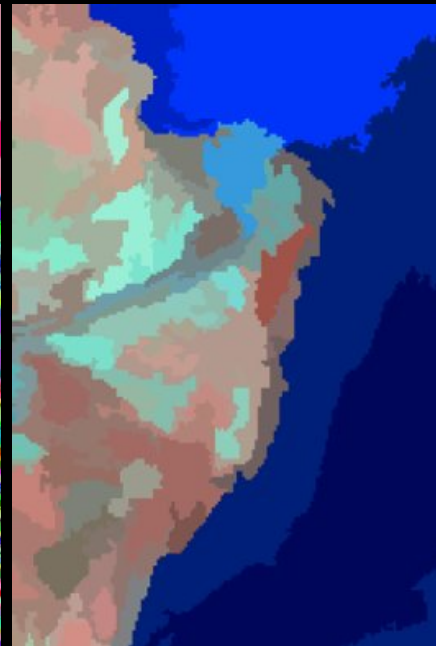
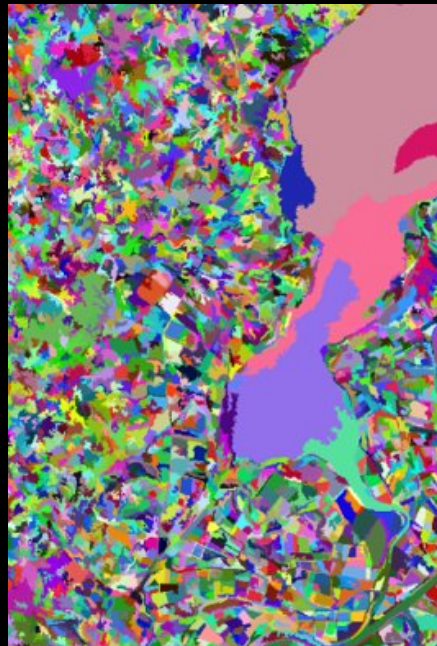
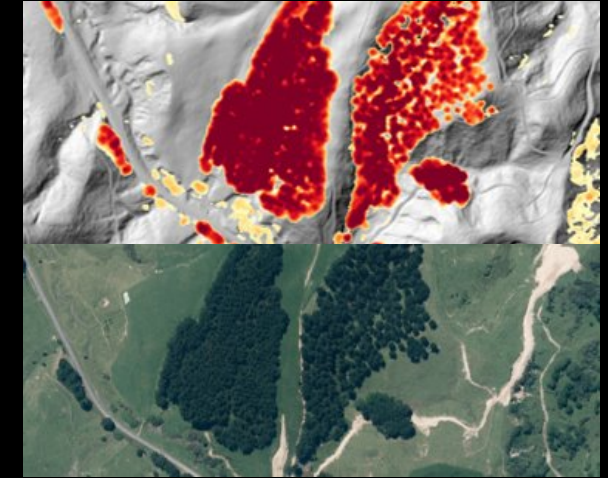
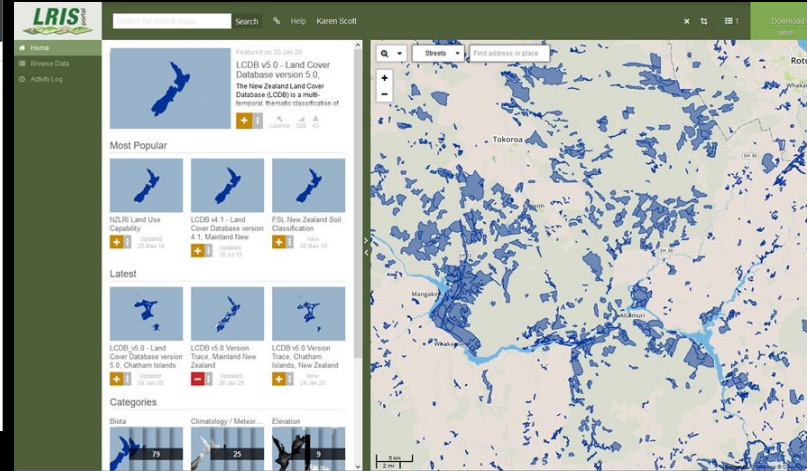
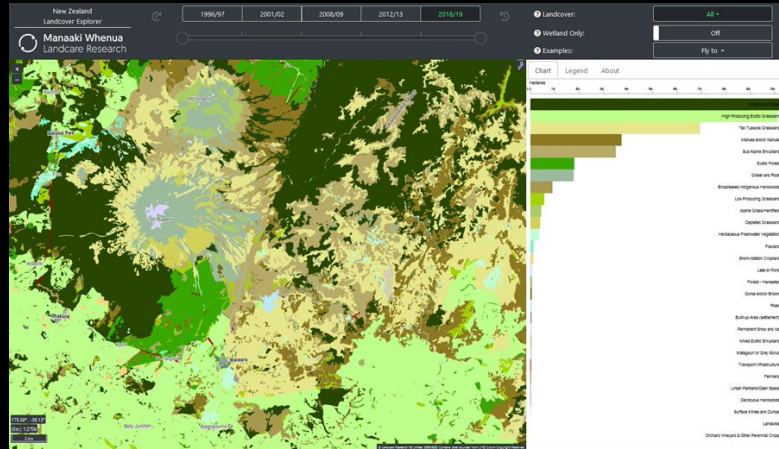
Starting an eResearch revolution with Deep Learning

Brent Martin and Aleksandra Pawlik
Manaaki Whenua Landcare Research

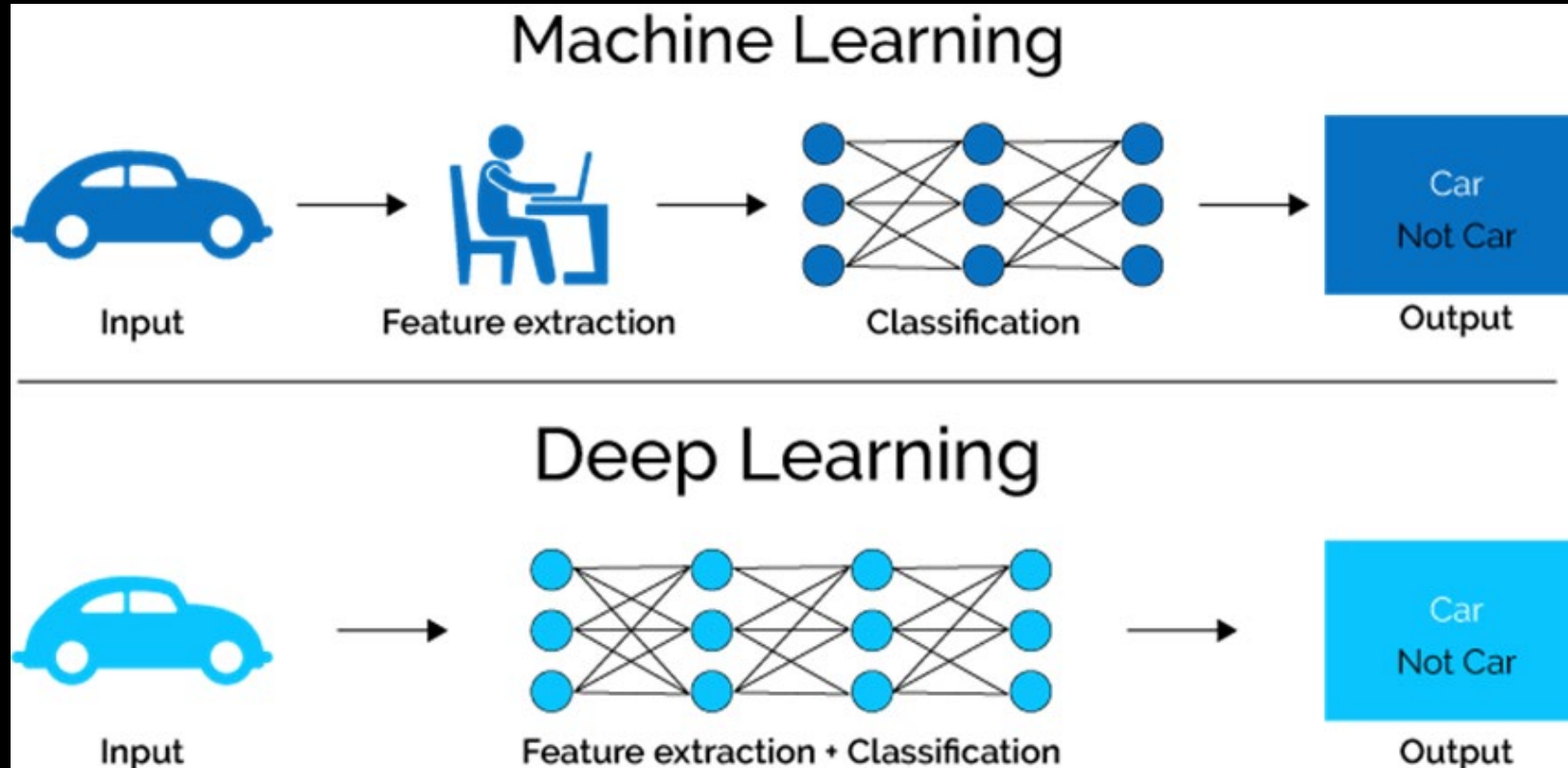
eResearch New Zealand 2021

10-12 February 2021, Wellington

Data deluge at Manaaki Whenua



Computer vision evolution – Deep Learning



**Imagenet 2010:
Error 28% (top 5)**

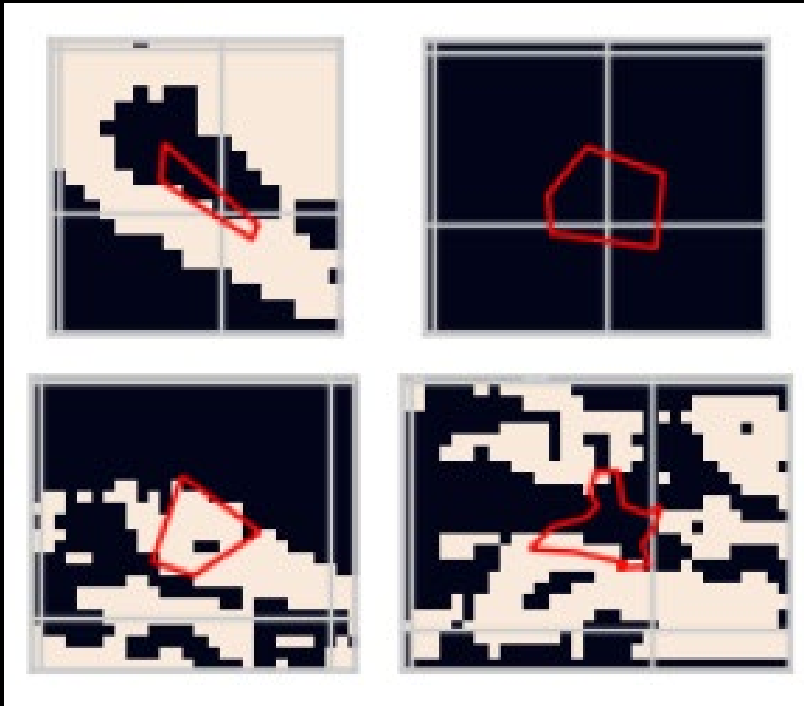
**Imagenet 2015:
Error 5% (top 5)**

Machine learning at Manaaki Whenua

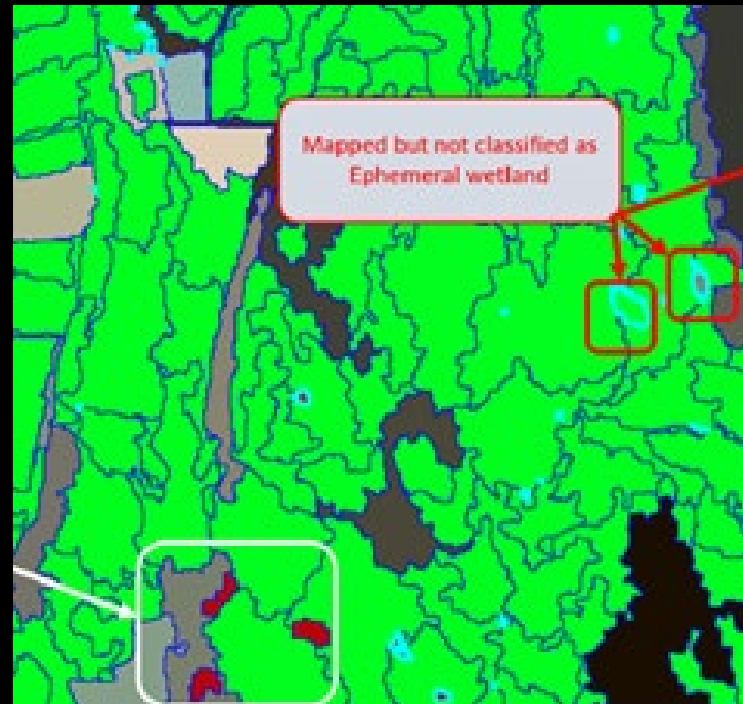


Object mapping (ephemeral wetlands)

Pixel classification: poor



Object classification: better



DL object detection: best?

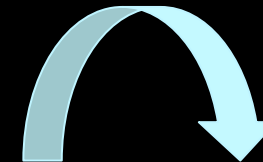


Increasing research impact at Manaaki Whenua



- Automate existing processes
- Extrapolate local ground survey data
- Detect change
- Discover new relationships and previously invisible patterns

3 months



1	LinkToVideoFile	TimeInterval	ReplicateInterval	ScientificName	TimeOfMax	MaxInterval
164	Site 10_GOPR0007	03:32:00	26	Paraperis collas	00:28:23	2
165	Site 10_GOPR0007	04:32:00	27	Paraperis collas	00:29:19	2
166	Site 10_GOPR0007	05:32:00	28	Paraperis collas	00:30:23	2
167	Site 10_GOPR0007	06:32:00	29	Paraperis collas	00:30:39	3
168	Site 10_GOPR0007	07:32:00	30	Paraperis collas	00:32:03	3
169	Site 11_GOPR0010	03:43:00	0	Paraperis collas	00:04:40	4
170	Site 11_GOPR0010	04:43:00	1	Paraperis collas	00:05:37	5
171	Site 11_GOPR0010	05:43:00	2	Paraperis collas	00:05:54	5
172	Site 11_GOPR0010	06:43:00	3	Paraperis collas	00:07:05	4
173	Site 11_GOPR0010	07:43:00	4	Paraperis collas	00:08:08	6

Beginnings – image recognition



Figure 1 Moth main species: Cebysa, Hyphantria, Lymantria, Mycetemera, Orgyia, Teia, Tyria, Utetheisa

Moths



Figure 2 Nyctemera amicus, cross, and annulata



Figure 3 Utetheisa lotrix and pulchelloides



Coprosmas



Fungi

Biosecure-ID (MBIE Smart Idea)

ACTUAL/PREDICTED	C-L	H-C	L-D	N-Am	N-X	N-An	O-T	T-A	T-J	U-L	U-P	RECALL
Cebysa leucotelus	100	0	0	0	0	0	0	0	0	0	0	1
Hyphantria cunea	0	90	0	0	0	0	0	0	0	0	0	1
Lymantria dispar	0	0	16	0	0	2	2	0	0	0	0	0.8
Nyctemera amicus	0	0	0	105	19	6	0	0	0	0	0	0.808
Nyctemera amicus x annulata	0	0	0	19	493	28	0	0	0	0	0	0.913
Nyctemera annulata	0	0	0	4	8	398	0	0	0	0	0	0.971
Orgyia thyellina	0	0	0	0	0	2	78	0	0	0	0	0.975
Teia anartoides	0	0	0	0	0	0	2	198	0	0	0	0.99
Tyria jacobaeae	0	0	0	0	0	0	0	1	279	0	0	0.996
Utetheisa lotrix	0	0	0	0	0	0	0	0	0	47	23	0.671
Utetheisa pulchelloides	0	0	0	0	0	0	0	0	0	31	239	0.885
PRECISION	1	1	1	0.82	0.95	0.91	1	1	1	0.6	0.9	

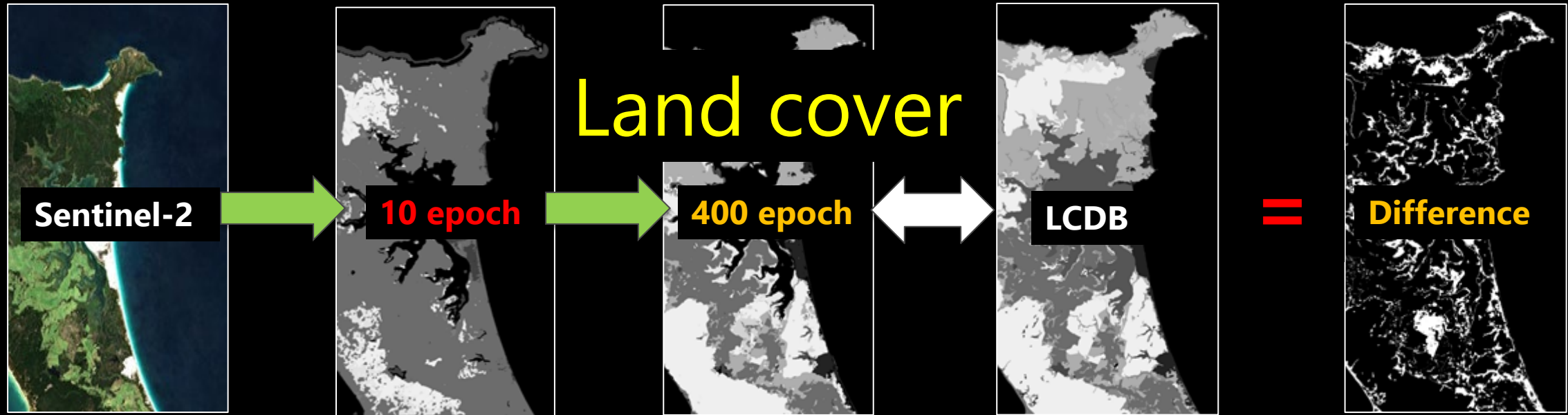
correct: 2043 (93.29%) incorrect: 147
Main species: 2181/9 (99.6%)

Moths confusion matrix

Species: 99.6%

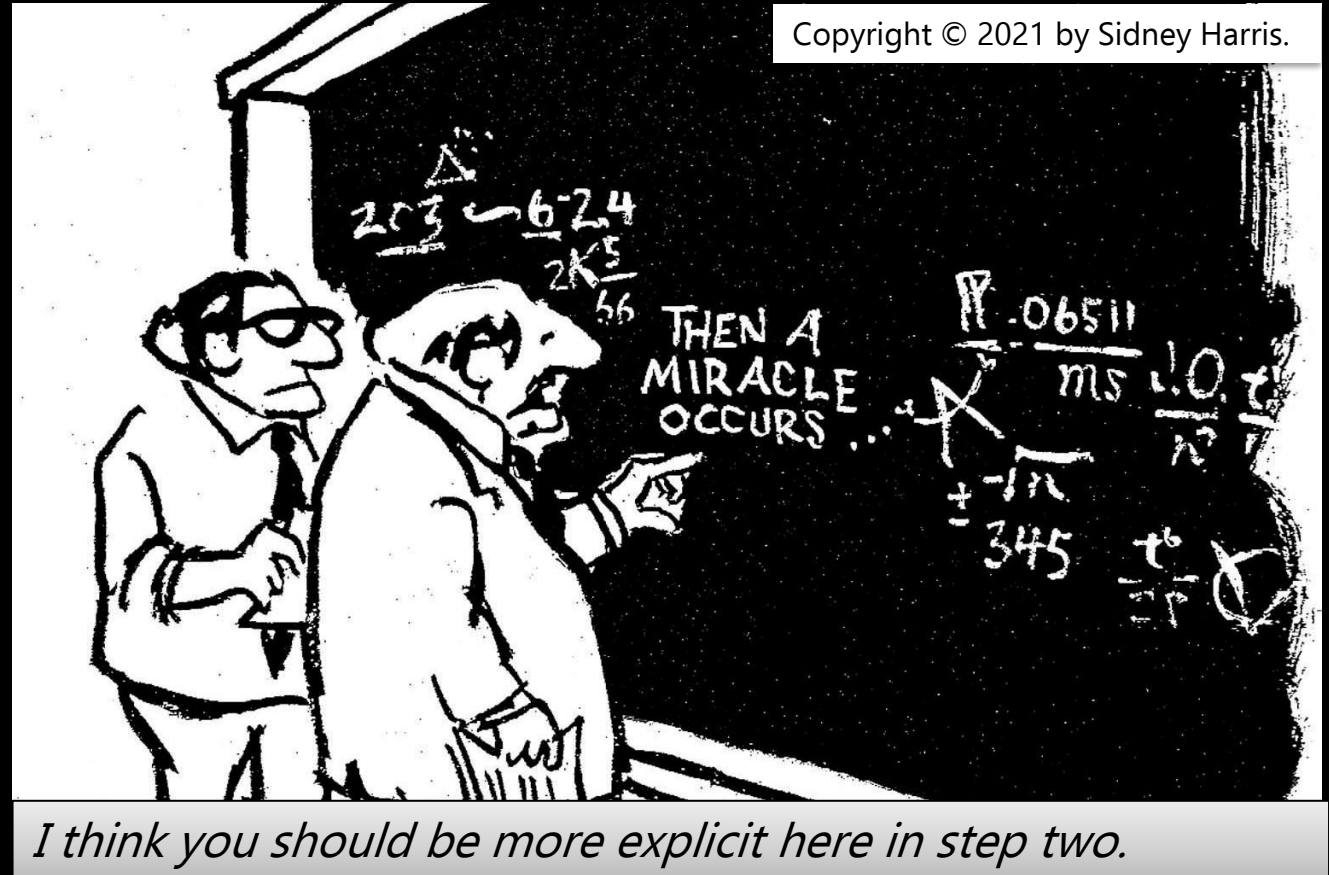
Subspecies or cross: 93.3%

Exploring – remote sensing



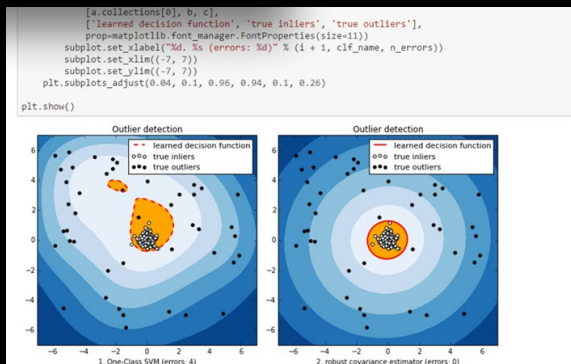
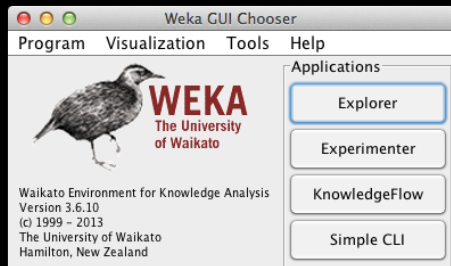
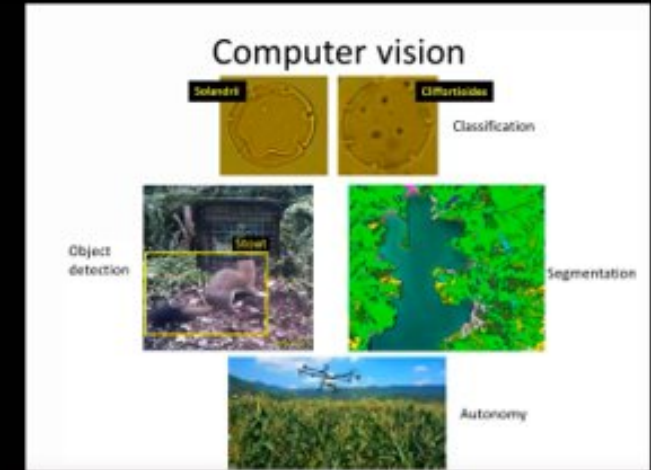
Key challenges

- Lack of awareness/trust
- "Black box" methodology
- Quality & quantity of training data
- Access to compute resources
- Gaps in skills set

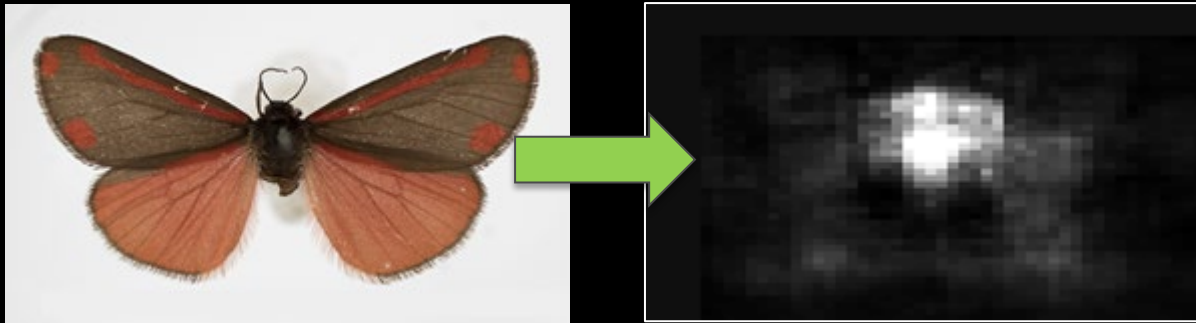
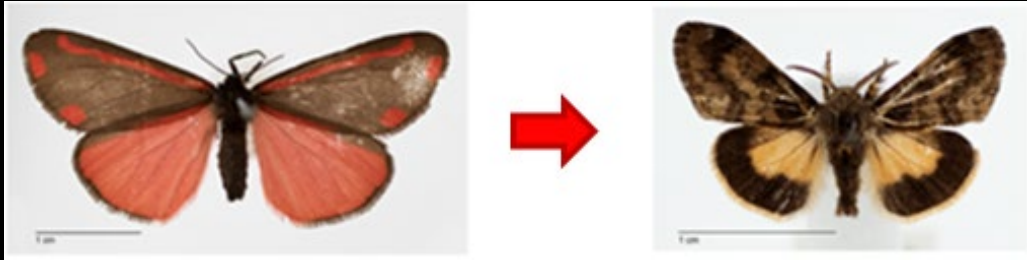


Solutions – people

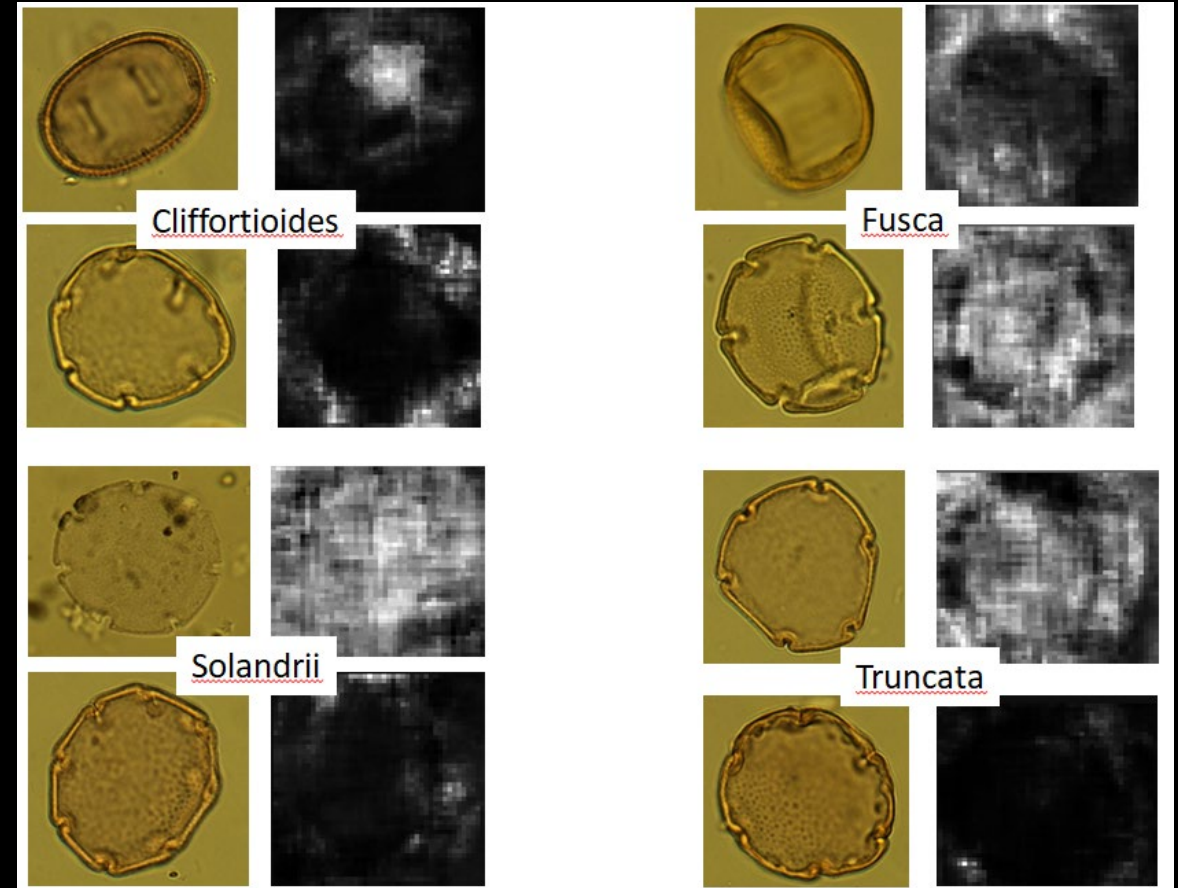
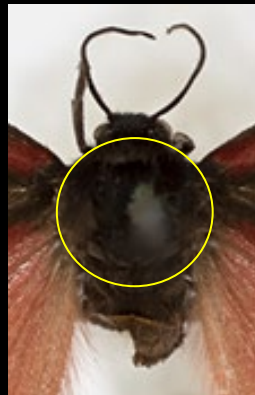
- Dissemination
- Training
- Resources
- Infrastructure



Solutions – black box: visualisation

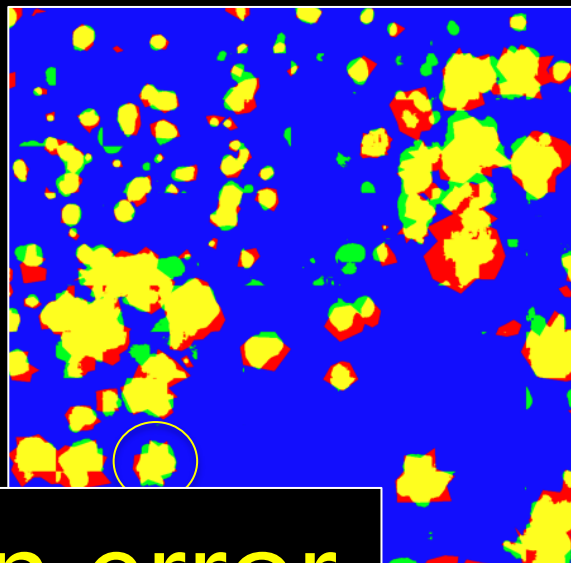


Explaining failure

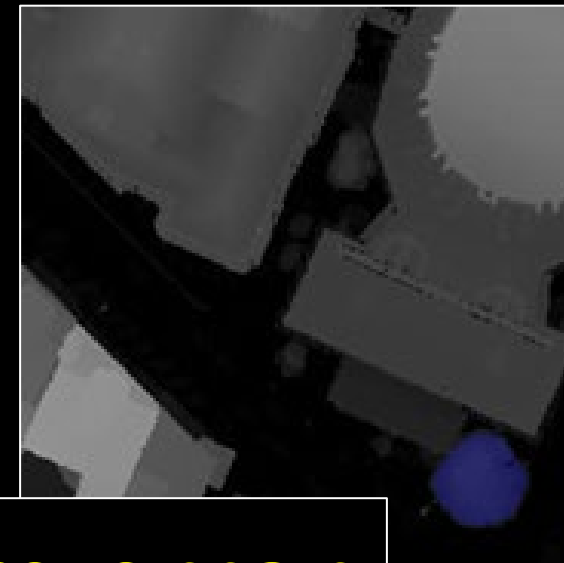
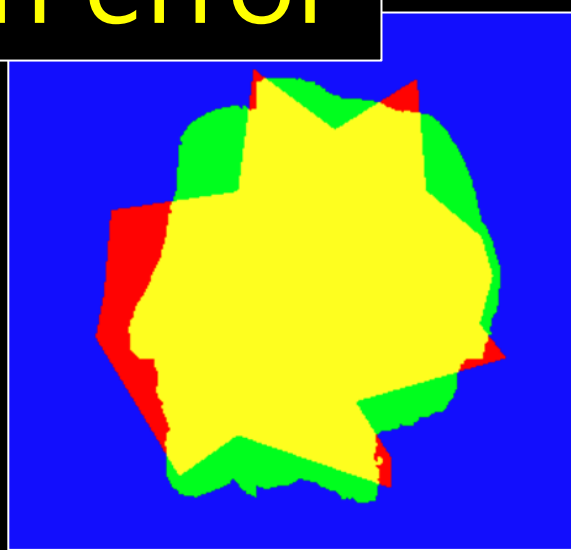


Explaining success!

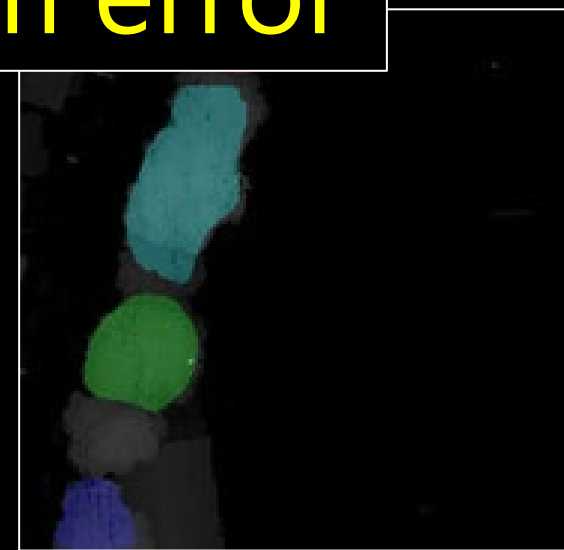
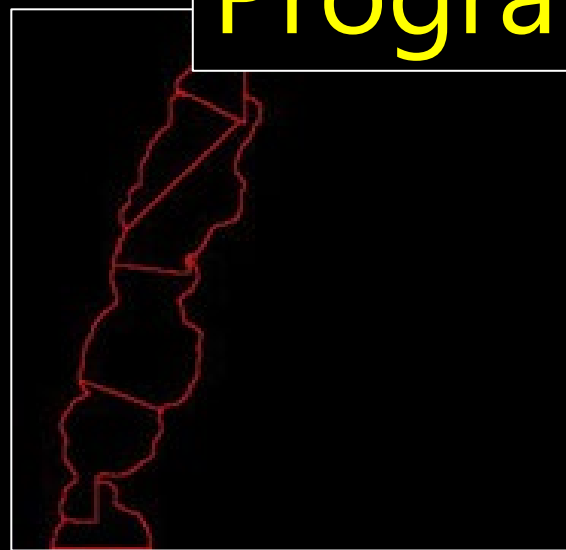
Data quality tolerance: bootstrapping



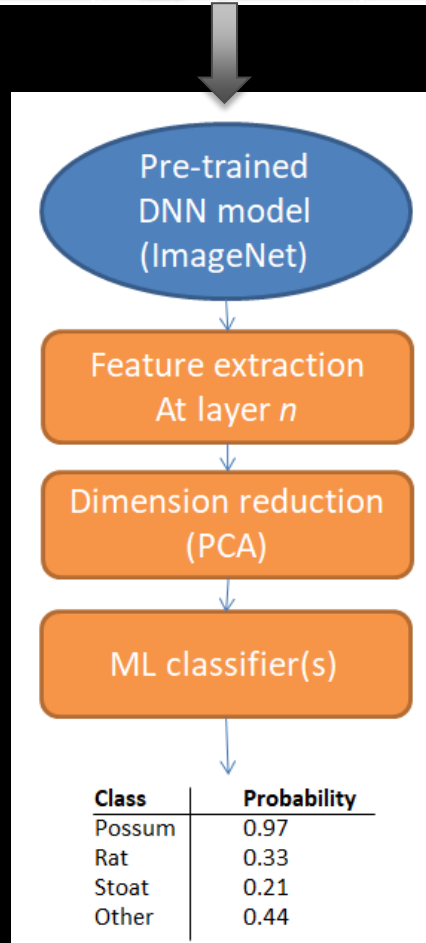
Human error



Program error



Data quantity tolerance: feature extraction



ACTUAL/PREDICTED	C-L	H-C	L-D	N-Am	N-X	N-An	O-T	T-A	T-J	U-L	U-P	RECALL
Cebysa leucotelus	100	0	0	0	0	0	0	0	0	0	0	1
Hyphantria cunea	0	90	0	0	0	0	0	0	0	0	0	1
Lymantria dispar	0	0	16	0	0	2	2	0	0	0	0	0.8
Nyctemera amicus	0	0	0	105	19	6	0	0	0	0	0	0.808
Nyctemera amicus x annulata	0	0	0	19	493	28	0	0	0	0	0	0.913
Nyctemera annulata	0	0	0	4	8	398	0	0	0	0	0	0.971
Orgyia thyellina	0	0	0	0	0	2	78	0	0	0	0	0.975
Teia anartoides	0	0	0	0	0	0	2	198	0	0	0	0.99
Tyria jacobaeae	0	0	0	0	0	0	0	1	279	0	0	0.996
Utetheisa lotrix	0	0	0	0	0	0	0	0	0	47	23	0.671
Utetheisa pulchelloides	0	0	0	0	0	0	0	0	0	31	239	0.885
PRECISION	1	1	1	0.82	0.95	0.91	1	1	1	0.6	0.9	

correct: 2043 (93.29%) incorrect: 147

Main species: 2181/9 (99.6%)

✓ Seconds to train on a laptop

✓ Better than transfer learning for tiny datasets



Compute resources: NeSI



MNIST on CPU (2017)
8 hours



MNIST on NeSI (2020)
2.5 minutes



MNIST on GPU (2017)
6 minutes

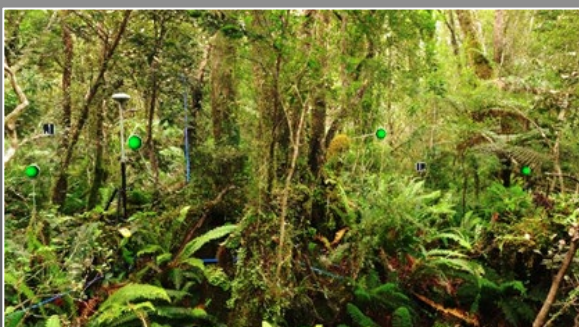
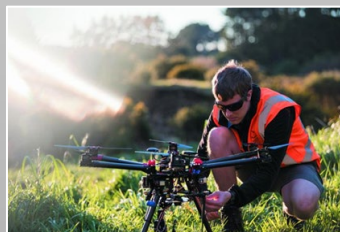
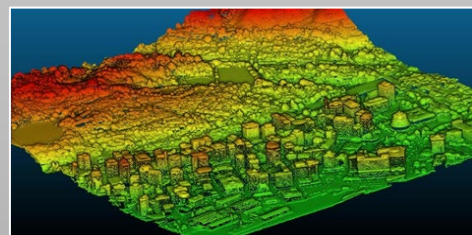


load_modules_fast.sh

```
1 |# loads the required modules for running Sheldon's fast transfer code
2 module load NeSI
3 module load Python/3.7.3-gimkl-2018b
4 module load CUDA/10.1.243
5 module load cuDNN/7.6.4.38-CUDA-10.1.243
```

Robust version control

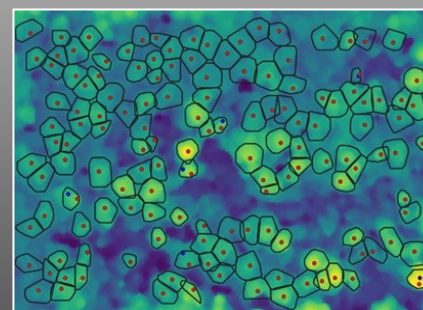
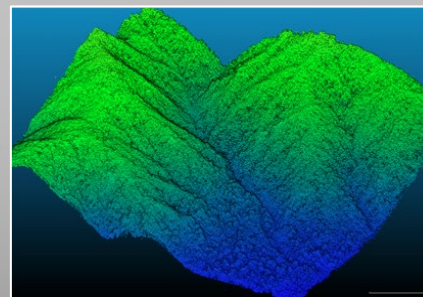
Putting it to work: remote tree identification



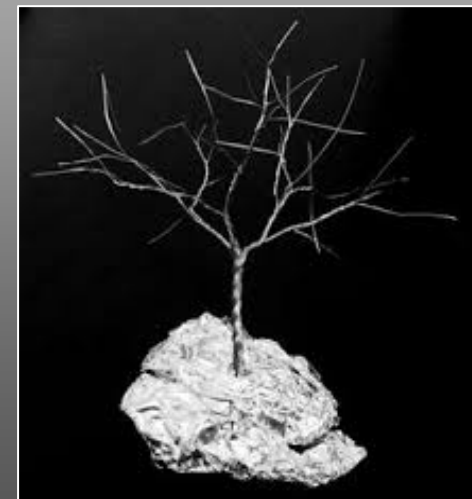
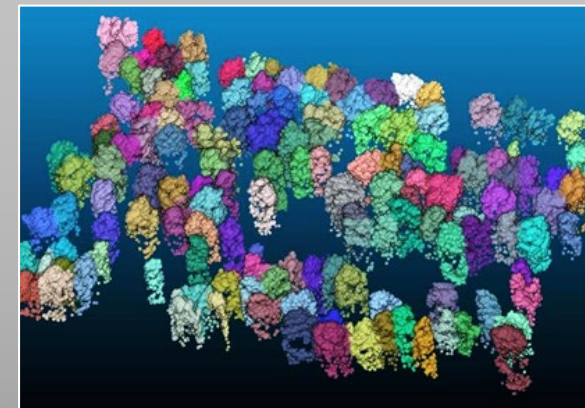
Prepare datasets



**NZ/Singapore
collaboration**



Extract training data



Train models



Looking forward

- Integration into our science
 - Special Interest Group
- "What we can offer" publication
- Funding
- Building up a platform / ecosystem
- External collaborations

Thank you