MET éireann

Centre for Research Training

Science Foundation Ireland sfi For what's next

Improving ECOCLIMAP Physiography Map using Machine Learning Techniques and Satellite Image data

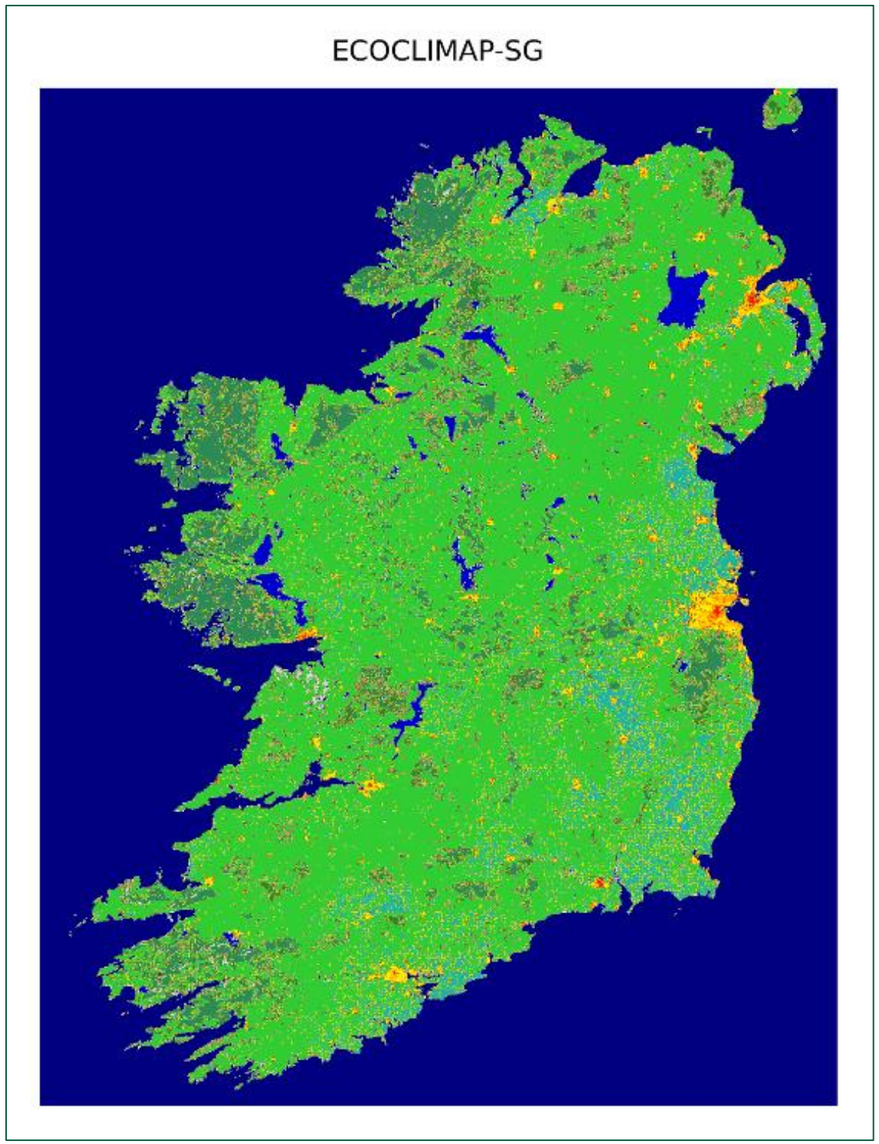
Eoin Walsh, Geoffrey Bessardon and Emily Gleeson

Thursday, October 1st 2020

UNIVERSITY OF LIMERICK OLLSCOIL LUIMNIGH | Bernal Institute

Project Motivation

- HARMONIE-AROME requires physiographic input – ECOCLIMAP (Météo-France).
- ECOCLIMAP can be improved– Icelandic Meteorological Office.
- Need for a universal mode of improvement.
- Need for method that can offer improved resolution.

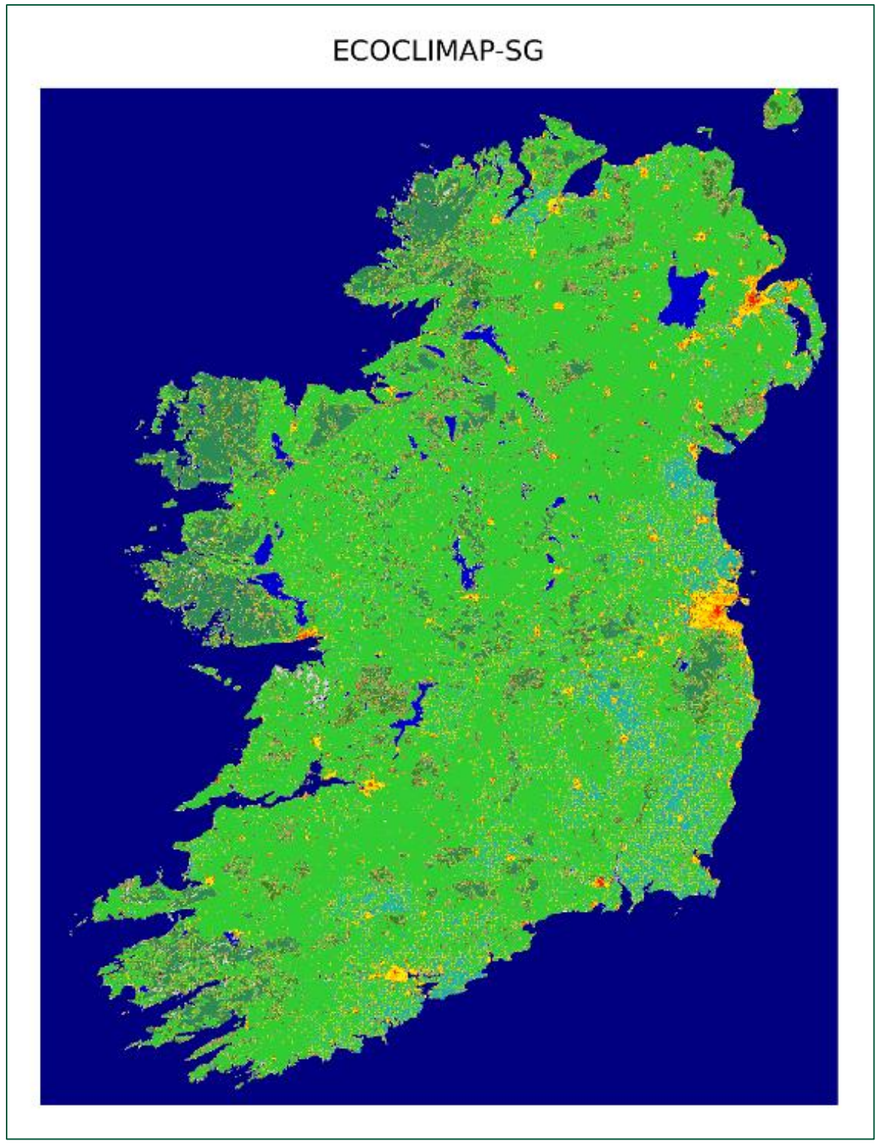


Project Motivation

1. Accuracy.

2. Resolution.

3. Universality.



Project Goals:

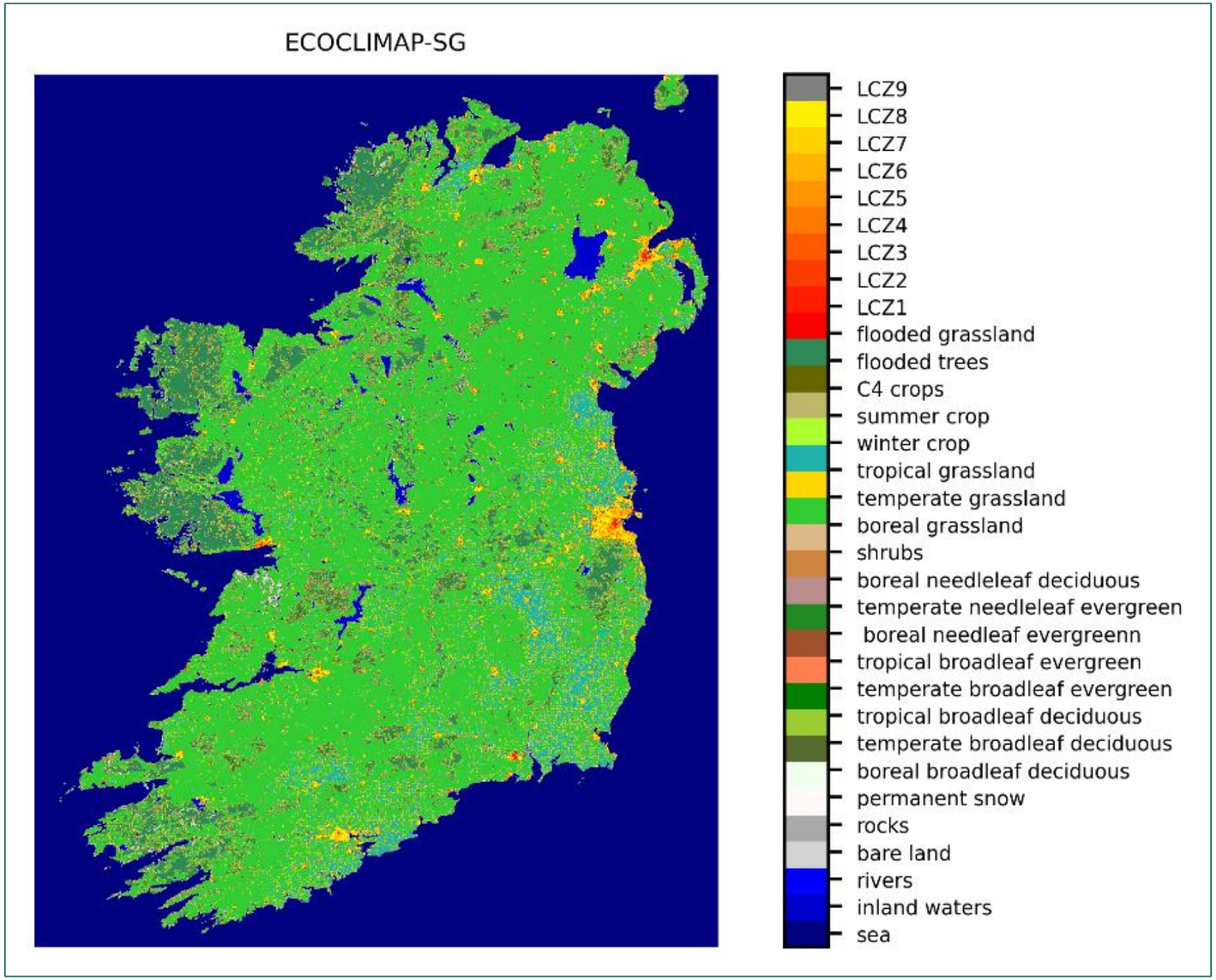
- Can ground-based image data and/or Satellite data be used:
    1. To increase the accuracy.
    2. To improve the resolution.
    3. For universal improvement.

Resources:

- Google Streetview API.
- CORINE Land Cover Map. (European Environmental Agency)
- Copernicus Open Access Hub for Sentinel-2 Data. (European Space Agency)

ECOCLIMAP-SG

- Global Landcover database, 1km resolution for 1st generation.
- ECOCLIMAP-SG has resolution of 300m.
- 33 landcover types in SG, 23 of which are present in Ireland.



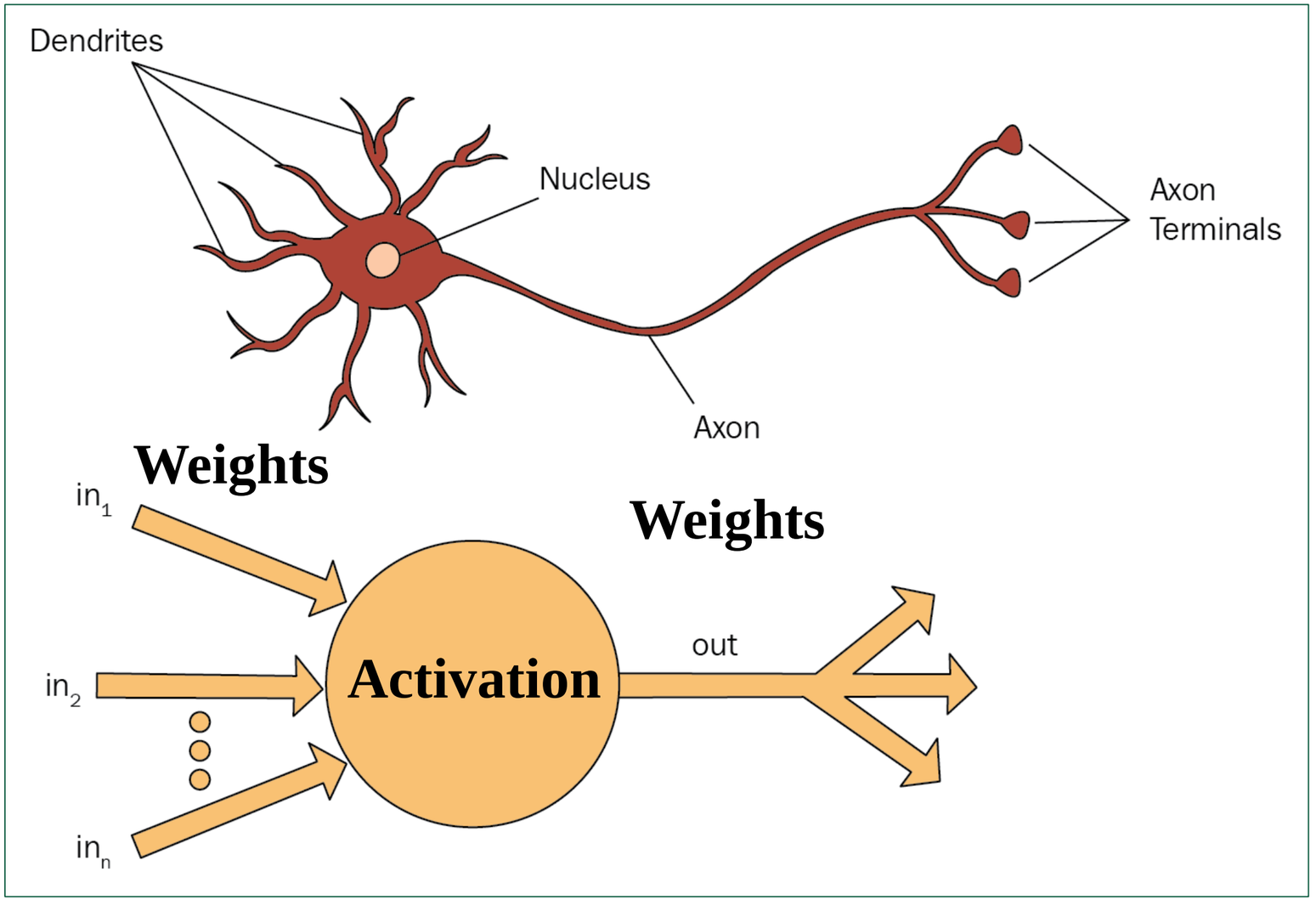
ECOCLIMAP-SG webpage:
https://opensource.umr-cnrm.fr/projects/ecoclimap-sg/

Machine Learning

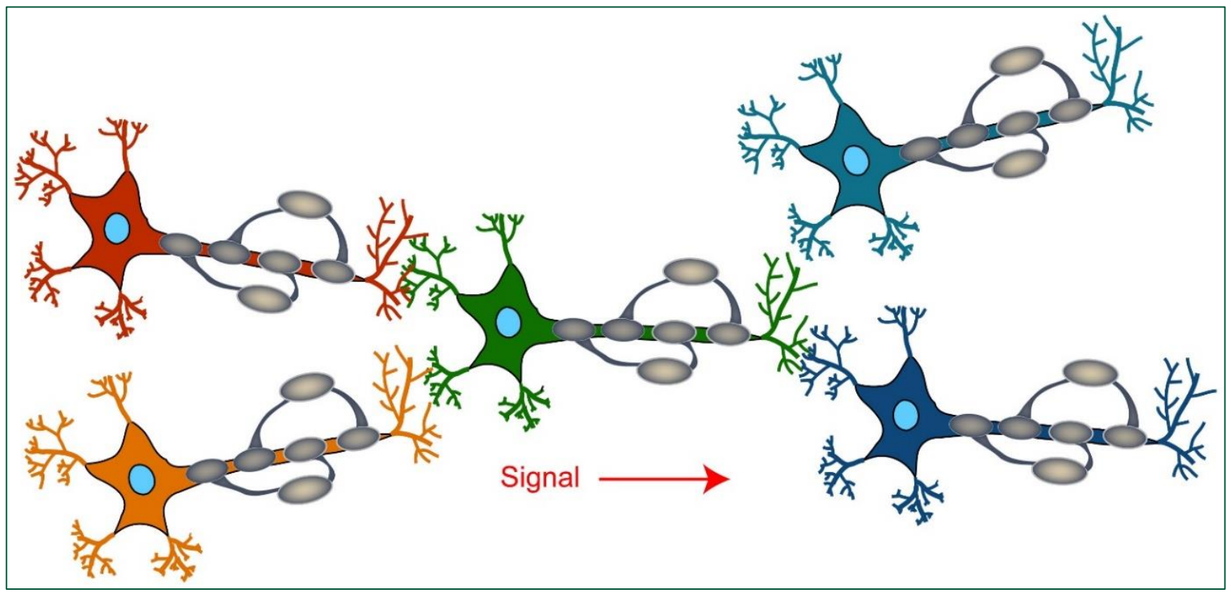
‘Machine Learning is the study of computer algorithms that improve automatically through experience’

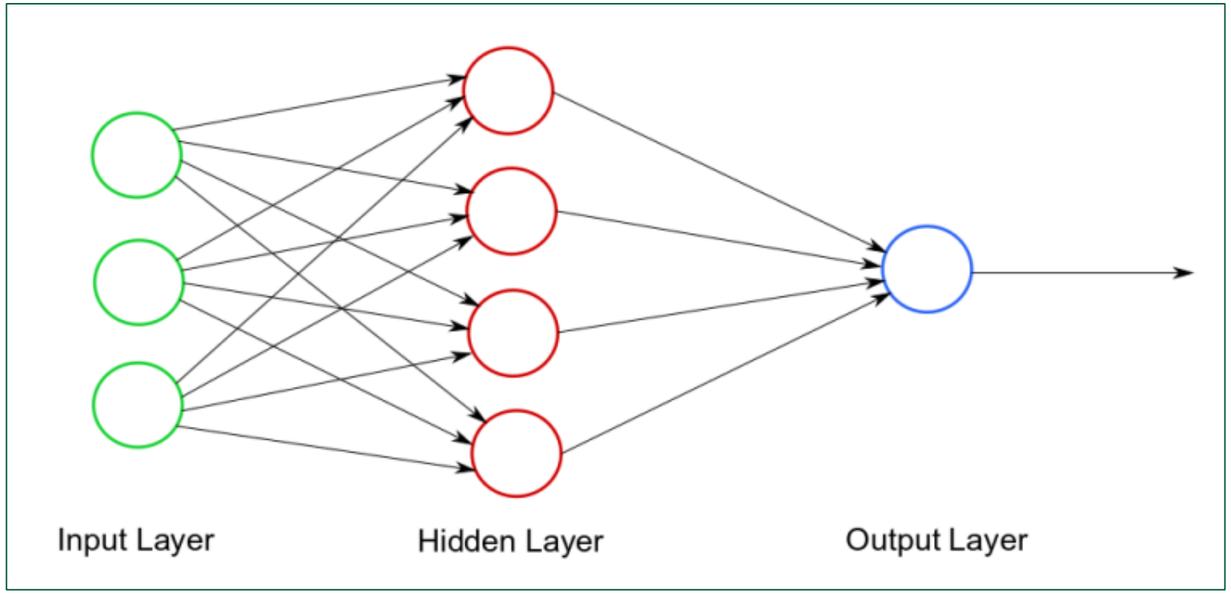
- Tom Mitchell, Machine Learning (1997).

Neural Networks



Neural Networks





Convolutional Neural Networks (CNNs)

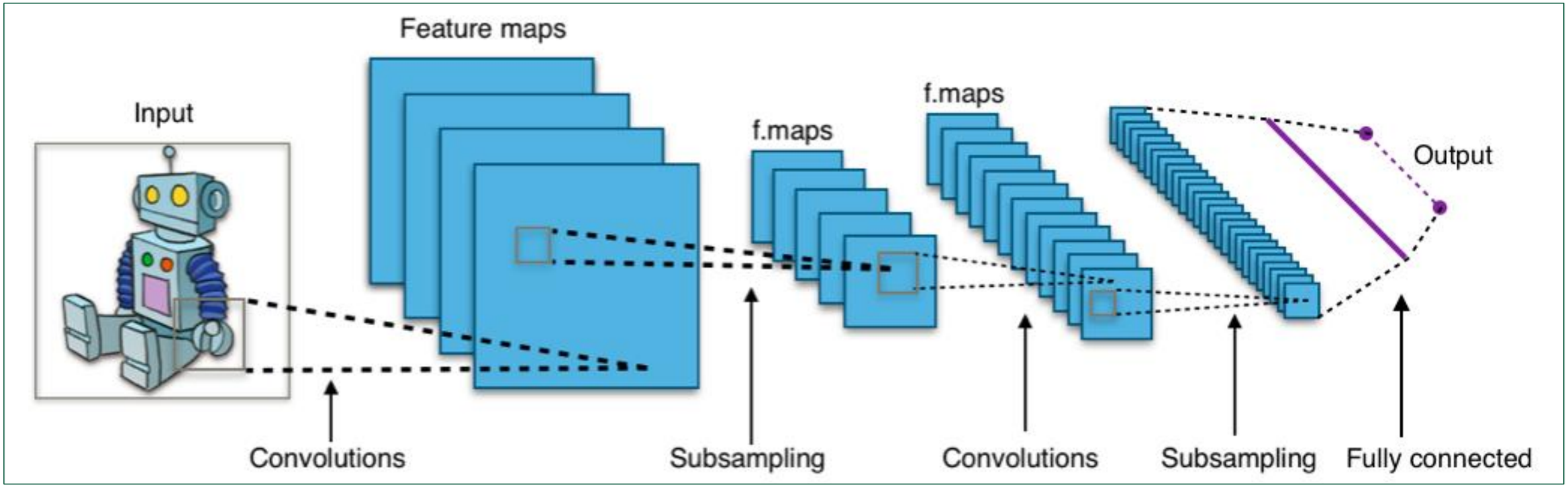


Image Source: https://medium.com/@himadrisankarchatterjee/a-basic-introduction-to-convolutional-neural-network-8e39019b27c4

CNN for Image Classification

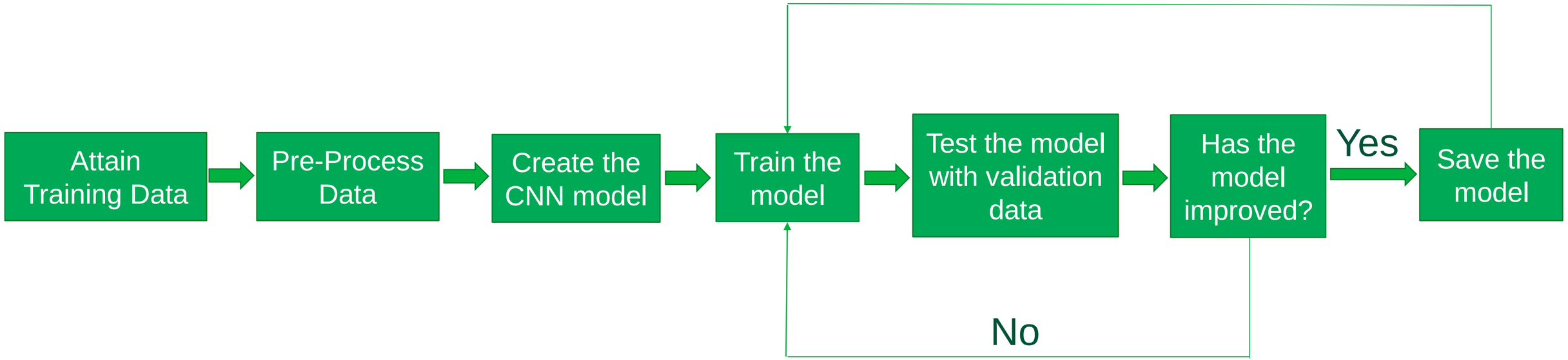




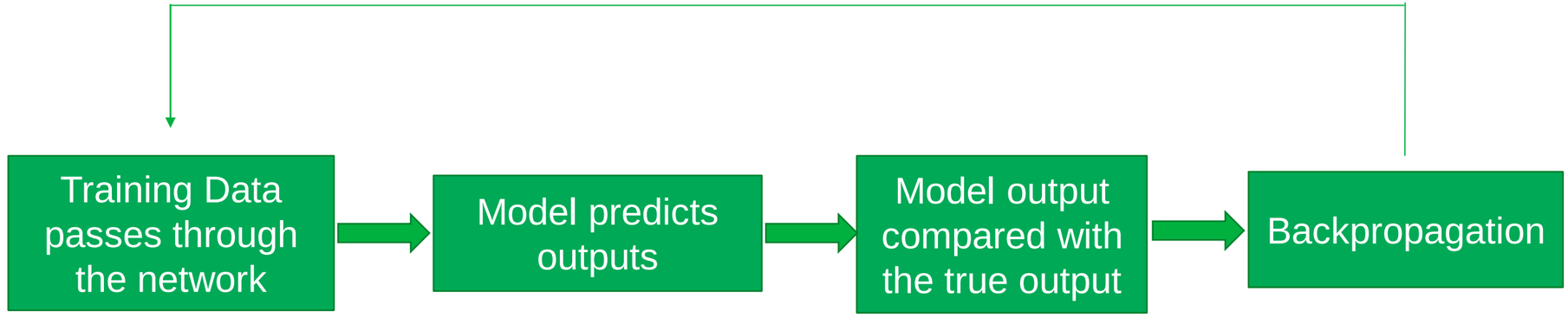


The process of a machine learning

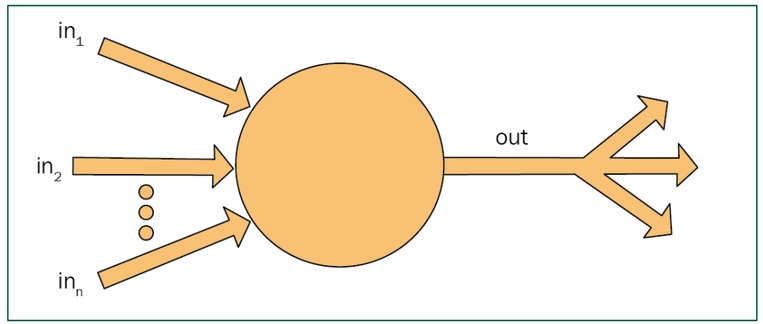
- Training Data – input data and corresponding correct output.
- Validation Data – Unseen Data to test the trained model.
- Epoch – One iteration of all of the training + validation data through the model.



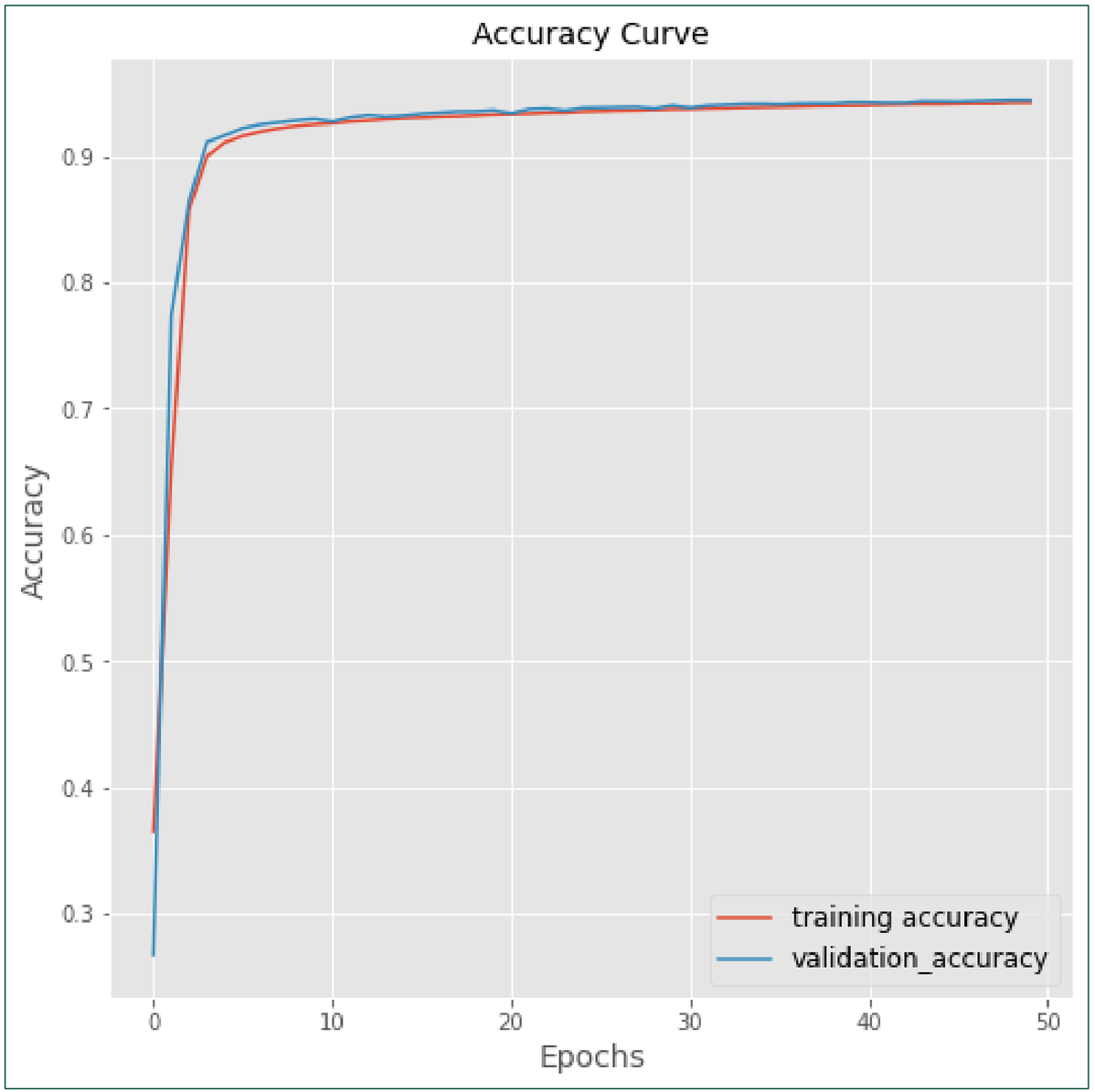
How does a model ‘learn’?



- Backpropagation:
  - The activations and weights of the network get adjusted appropriately, so as to reduce the error value.



How does a model ‘learn’?



CNN for Image Segmentation

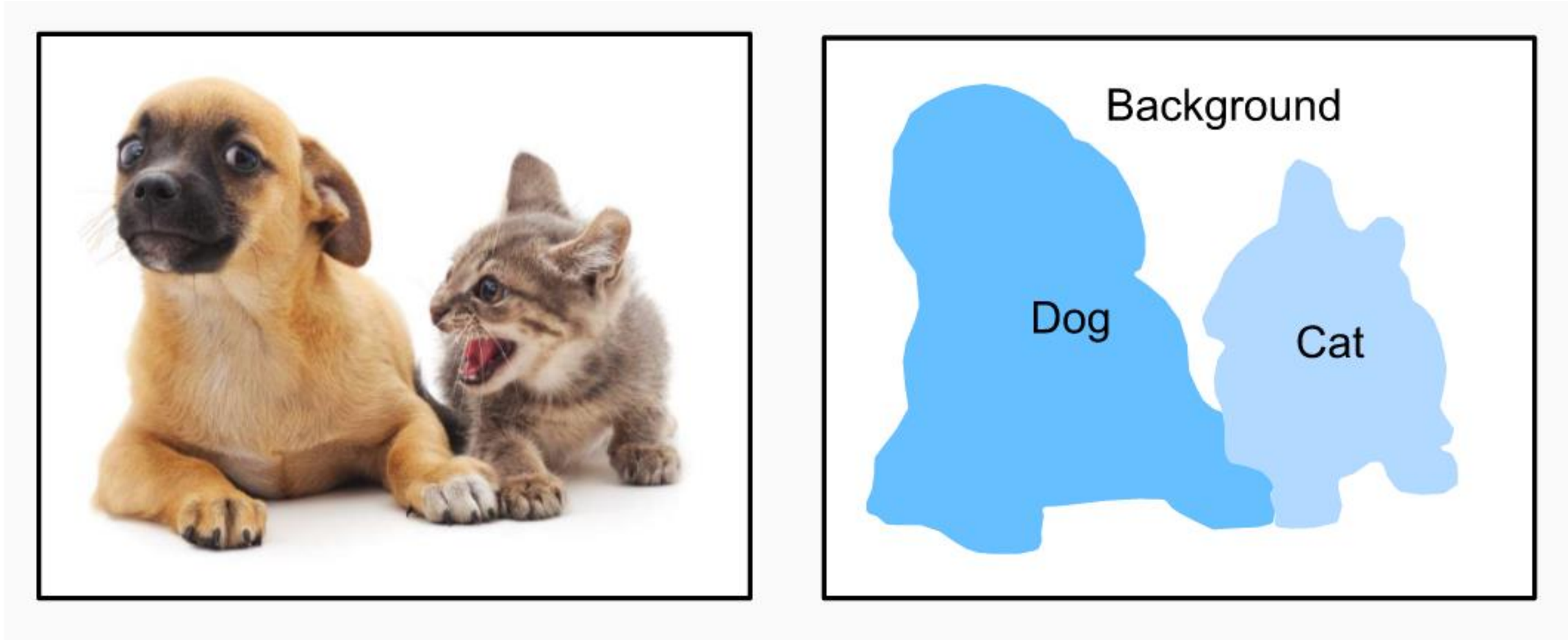


Image Source: https://d2l.ai/chapter_computer-vision/semantic-segmentation-and-dataset.html

Transfer Learning 1

- A model trained for one task is re-purposed for use on a second related task.
- Re-purposing a CNN Classifier.





Transfer Learning 2

- A model trained for one task is re-purposed for use on a second related task.
- Re-purposing CNN Classifier as a segmentation model.

Horse

Images Source: https://www.learnopencv.com/pytorch-for-beginners-semantic-segmentation-using-torchvision/

Transfer Learning 2

- A model trained for one task is re-purposed for use on a second related task.
- Re-purposing CNN Classifier as a segmentation model.

Images Source: https://www.learnopencv.com/pytorch-for-beginners-semantic-segmentation-using-torchvision/

12 Week Synopsis

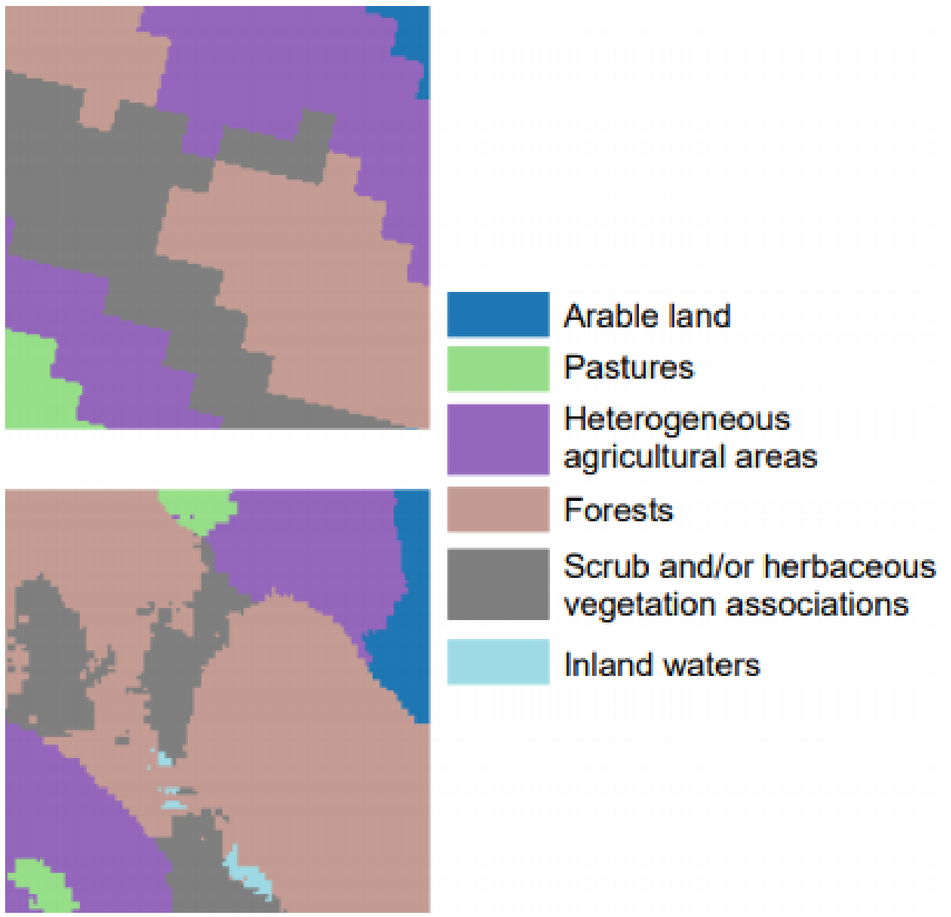
- First 4 weeks:
  - ➢ Inv
  - ➢ Go
  - ➢ CN d no bui

12 Week Synopsis

12 week Synopsis

- Remaining 8 weeks:
  - ➢Investigation using Satellite Imagery.
  - ➢Sentinel-2 Images plus Corine Landcover Images.
    - ➢2 Segmentation algorithms.
      - ➢5 label algorithm.
      - ➢15 label algorithm.

Improving CORINE Land Cover map in Estonia



The segmentation model shows a high 91.4% pixel-level accuracy on the first classification level with 75.8% and 59.7% on the second and third levels (Table 3). An example of segmentation model results and comparison to validation data can be seen on Figures 3 and 4.

Source:

Ulmas, Priit; Liiv, Innar; “Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification”, March 2020.

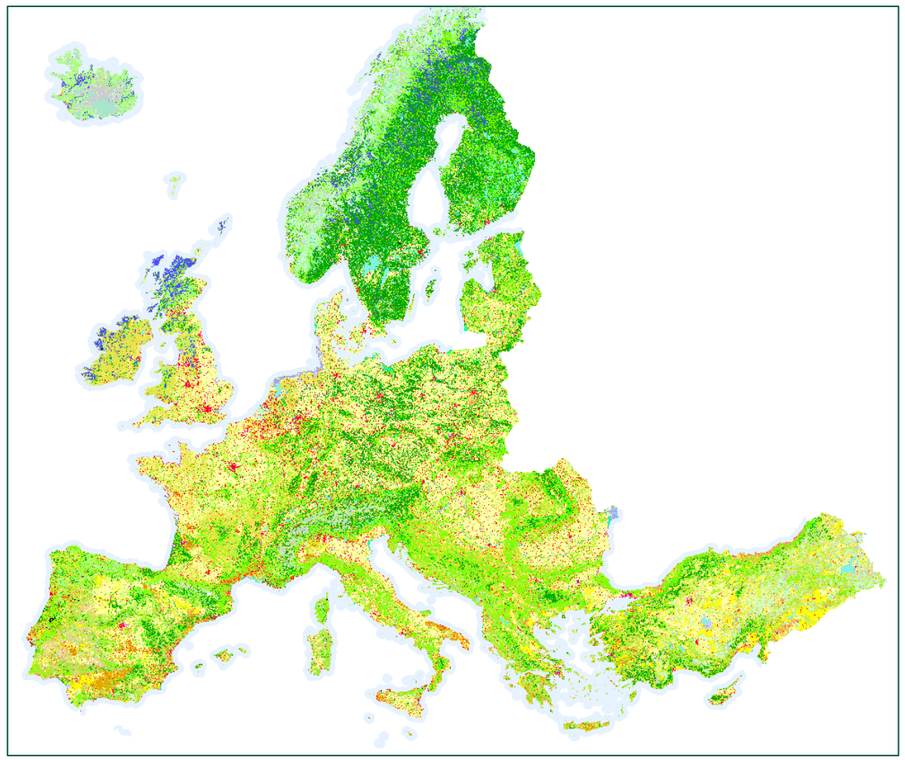
Datasets Used

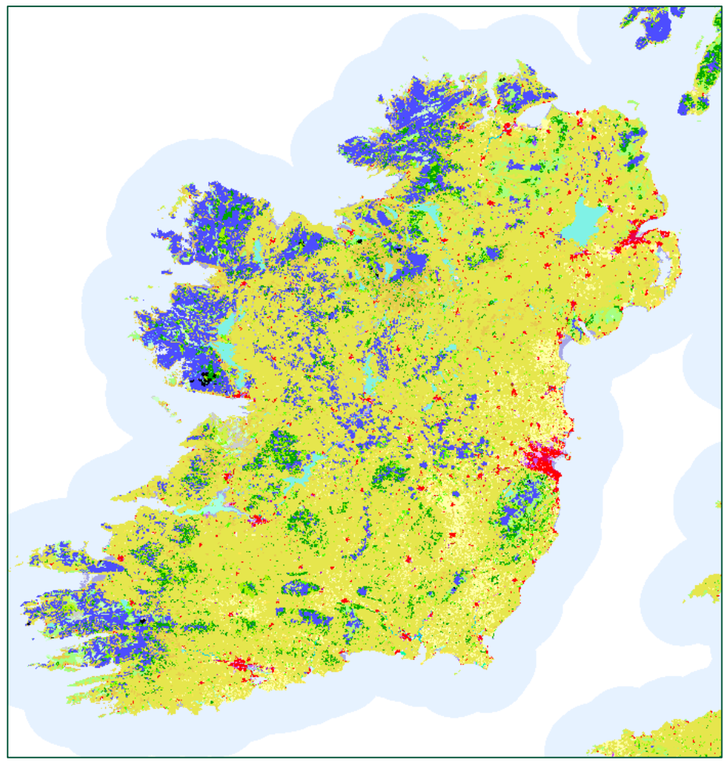
- BigEarthNet
  - ➢ ~ 600,000 sentinel segments ($1.2km^2$ each), appended with cover labels.

Website: bigearth.net

Datasets Used

- Corine Land Cover Map.
  - ➢ Split into Primary, Secondary and Tertiary Labels (5, 15 and 44 labels).
  - ➢100 metre resolution





Website: https://land.copernicus.eu/pan-european/corine-land-cover

Datasets Used

- Sentinel-2 Satellite Images.
  - ➢Obtained from Copernicus Open Access Hub.
  - ➢10 metre resolution.

Donegal

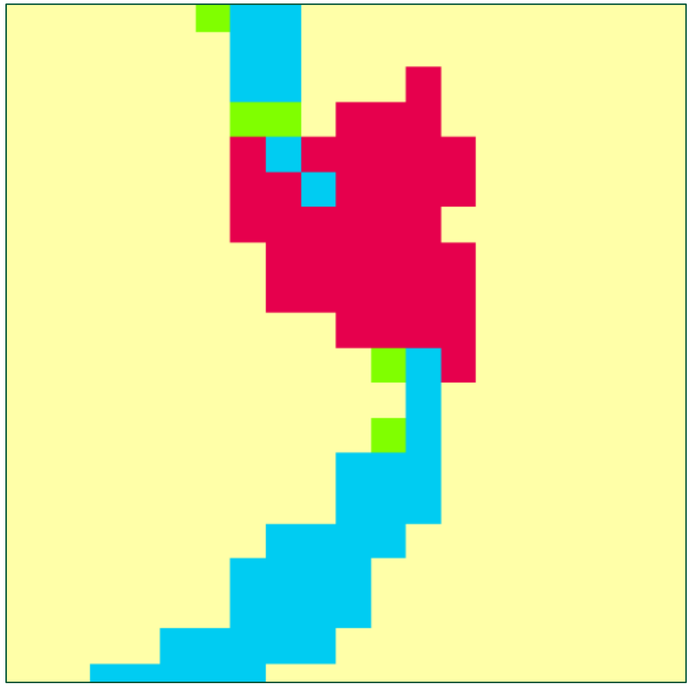
Dublin

Website: https://scihub.copernicus.eu/dhus/#/home

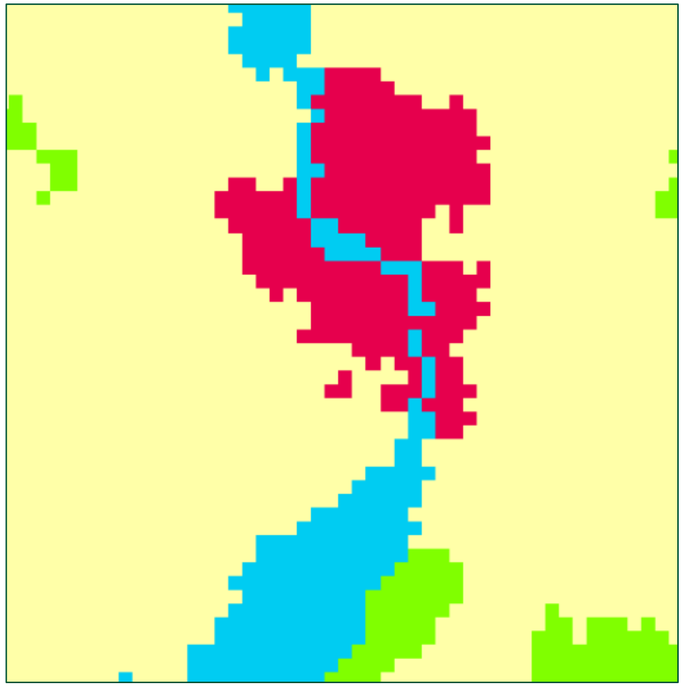
Why use Corine as training output?

- Corine (100m) has a better resolution than ECOCLIMAP-SG (300m)
- Corine considered to be 85% accurate. [1]

Sentinel-2



ECOCLIMAP-SG

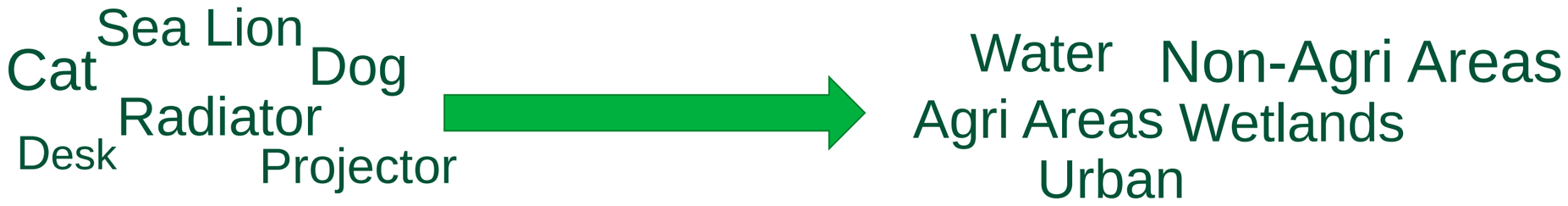


CORINE

[1] CORINE LAND COVER 2012 FINAL VALIDATION REPORT
Link: https://land.copernicus.eu/user-corner/technical-library/clc-2012-validation-report-1

The Model

- ‘Resnet-50’, pre-trained on the ImageNet dataset.
- Transfer Learning..



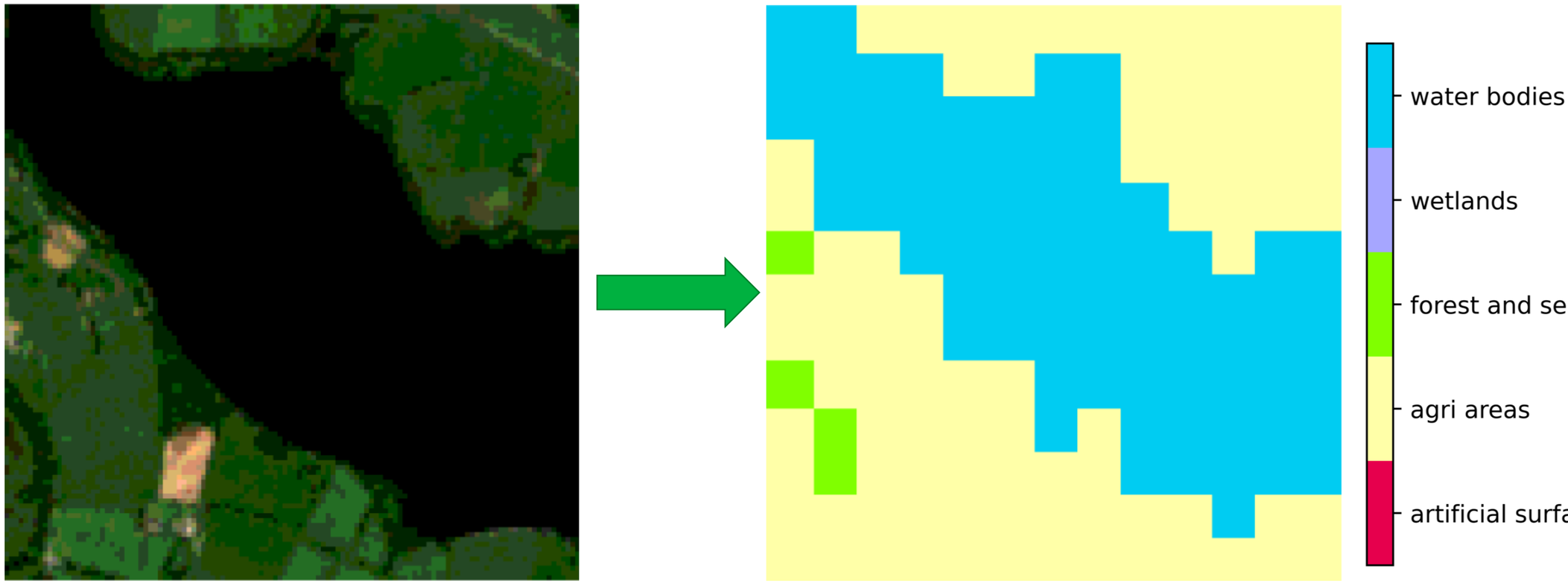
The Model

- Classifier Model re-purposed as a segmentation algorithm.

waters, agri-areas.

The Model

- Classifier Model re-purposed as a segmentation algorithm.



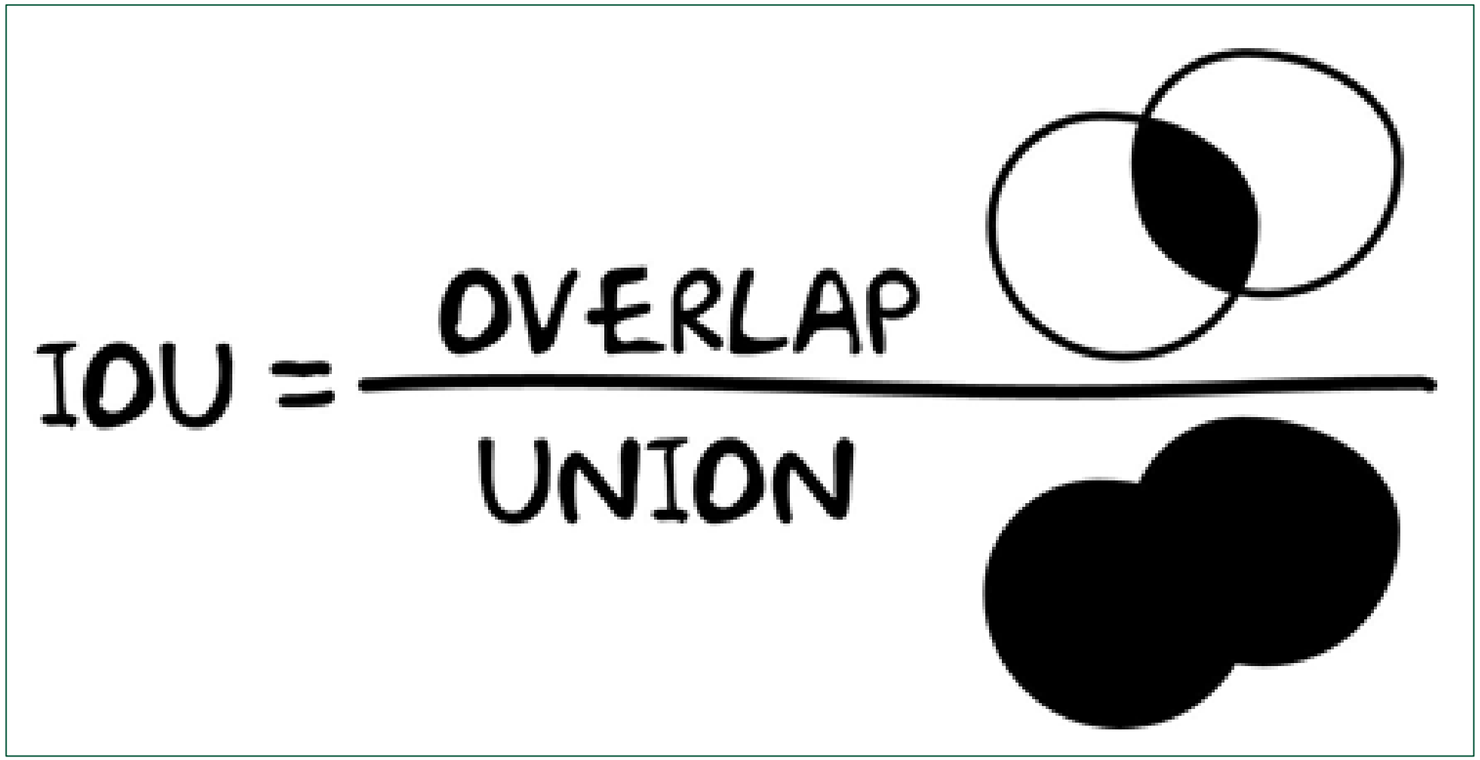
Analysing Model results

- Compare ECOCLIMAP and Model Prediction with best map we have – CORINE.

- Intersection over Union to compare accuracy of categories.

- Visual Inspection to demonstrate better accuracy.

Intersection Over Union (IOU)



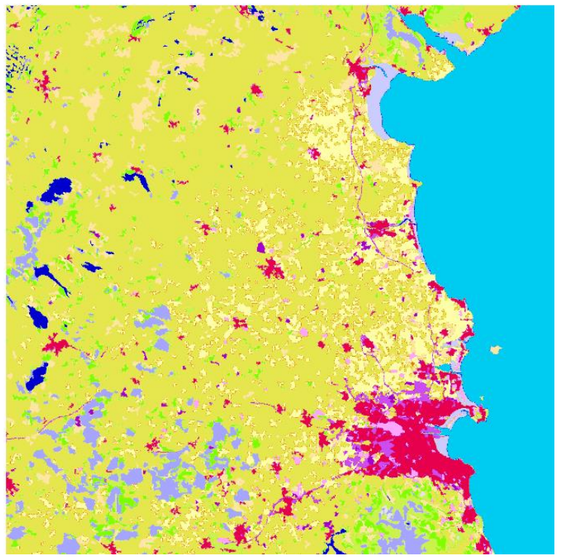
- IOU of 1 – prediction completely correct.
- IOU of 0 – prediction completely incorrect.

Primary Training Data

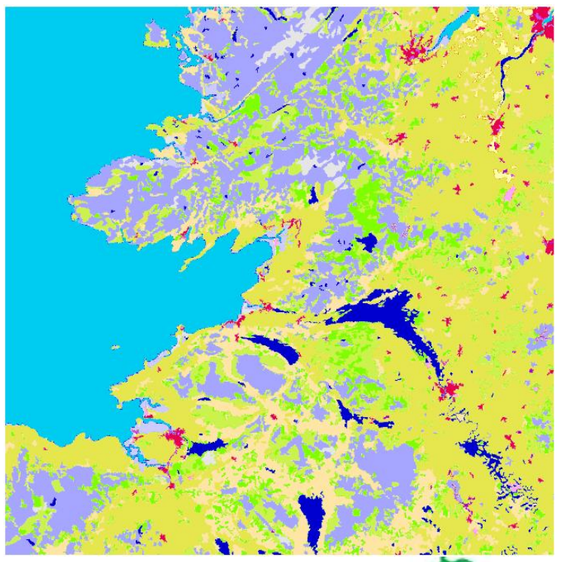
Sentinel-2 Segments

CORINE Segments

Dublin

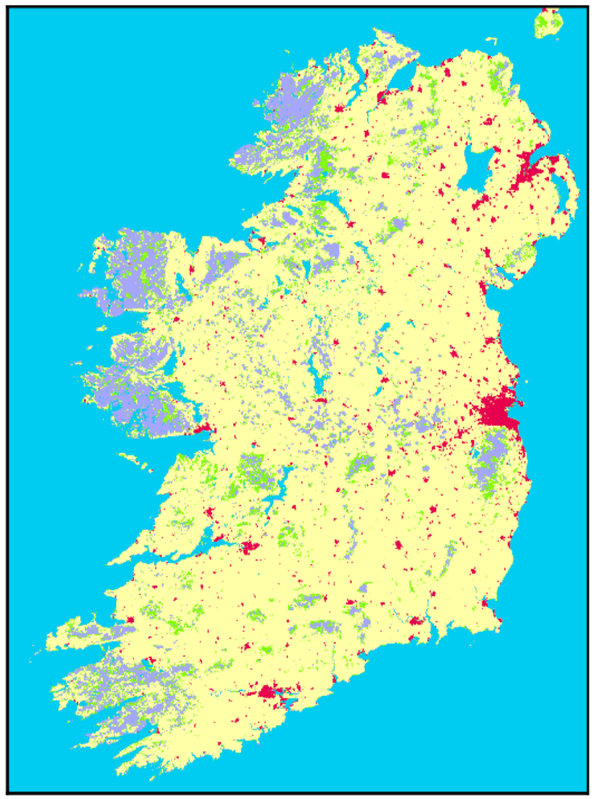


Donegal



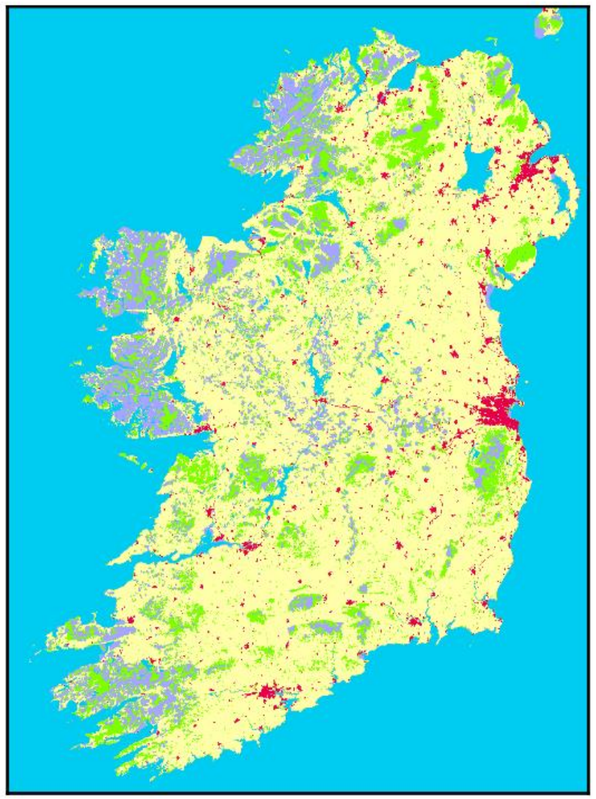
Primary Results

ECOCLIMAP

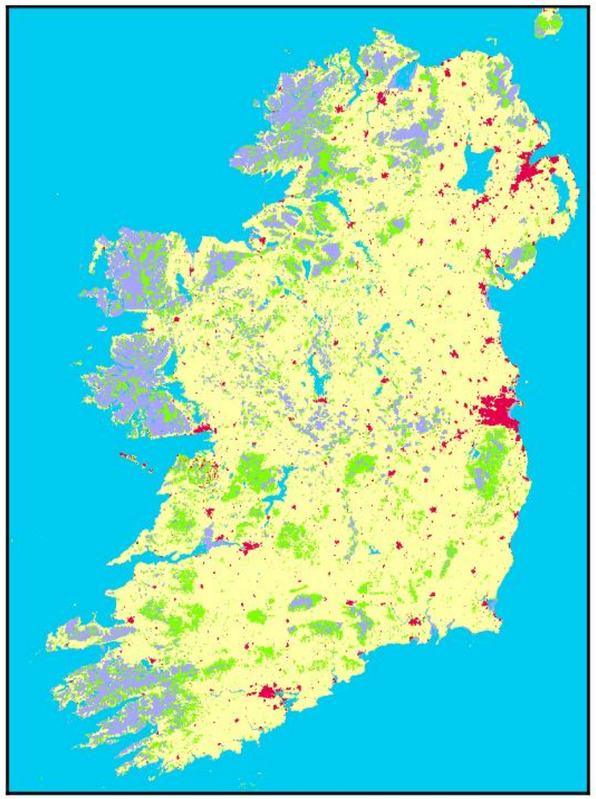


89.8%

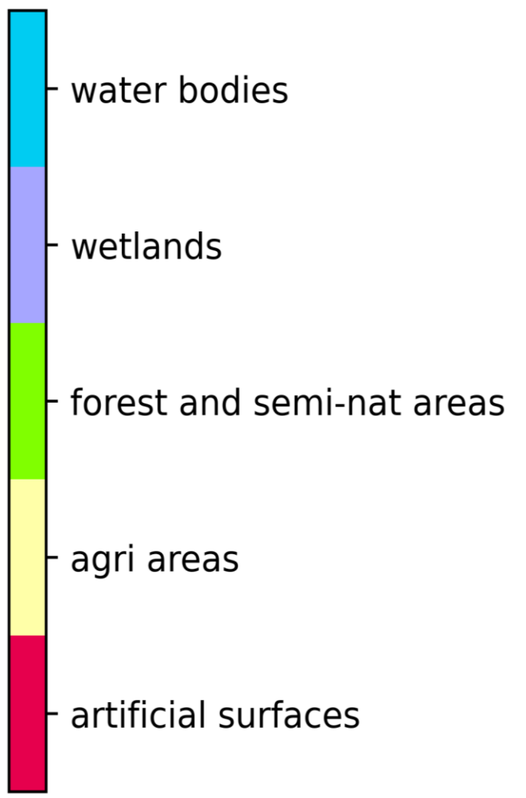
CORINE



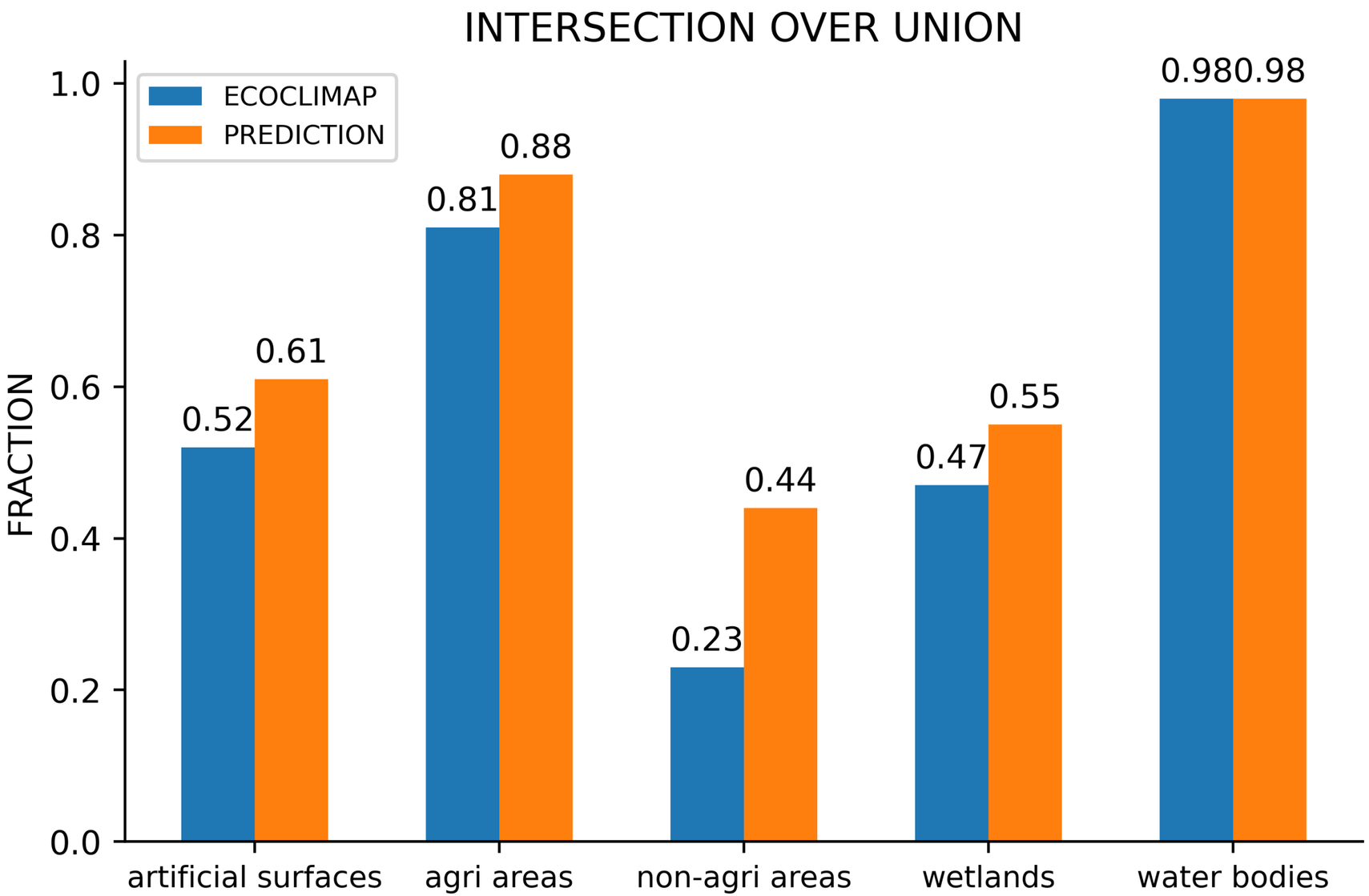
MODEL PREDICTION



92.6%

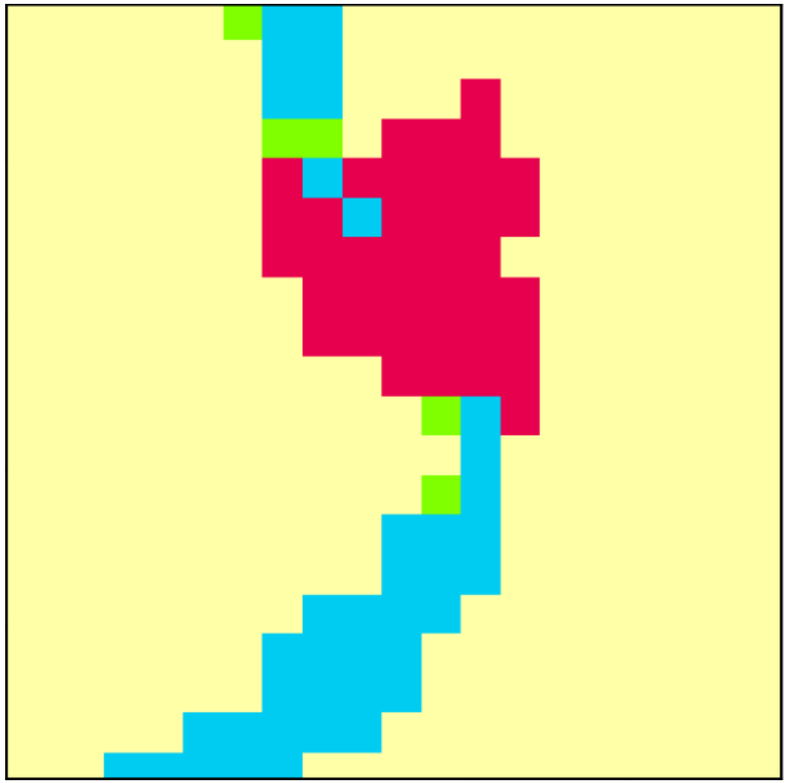


Primary Results - Accuracy

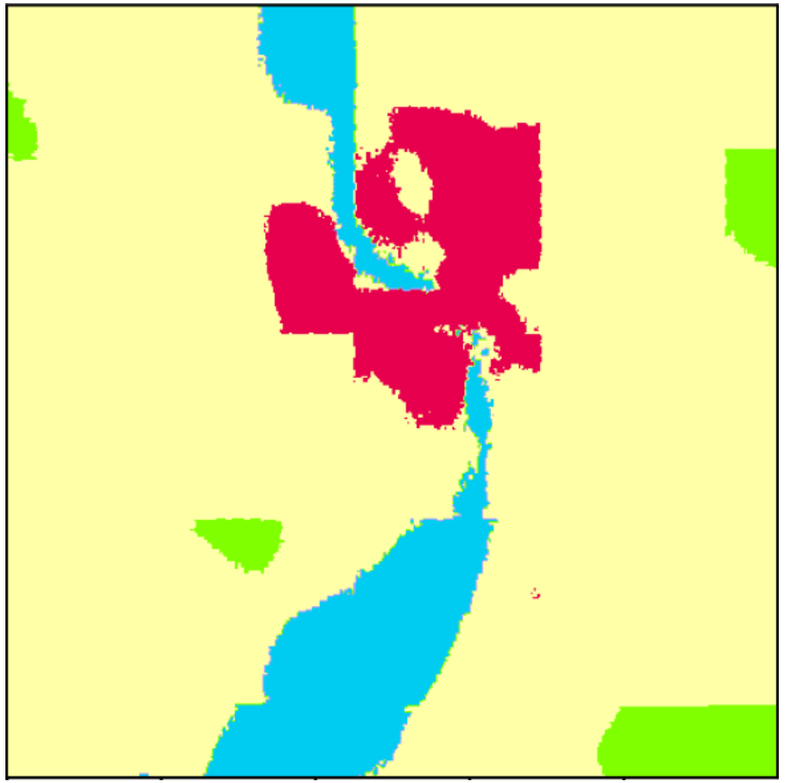


Primary Results - Resolution

Sentinel-2

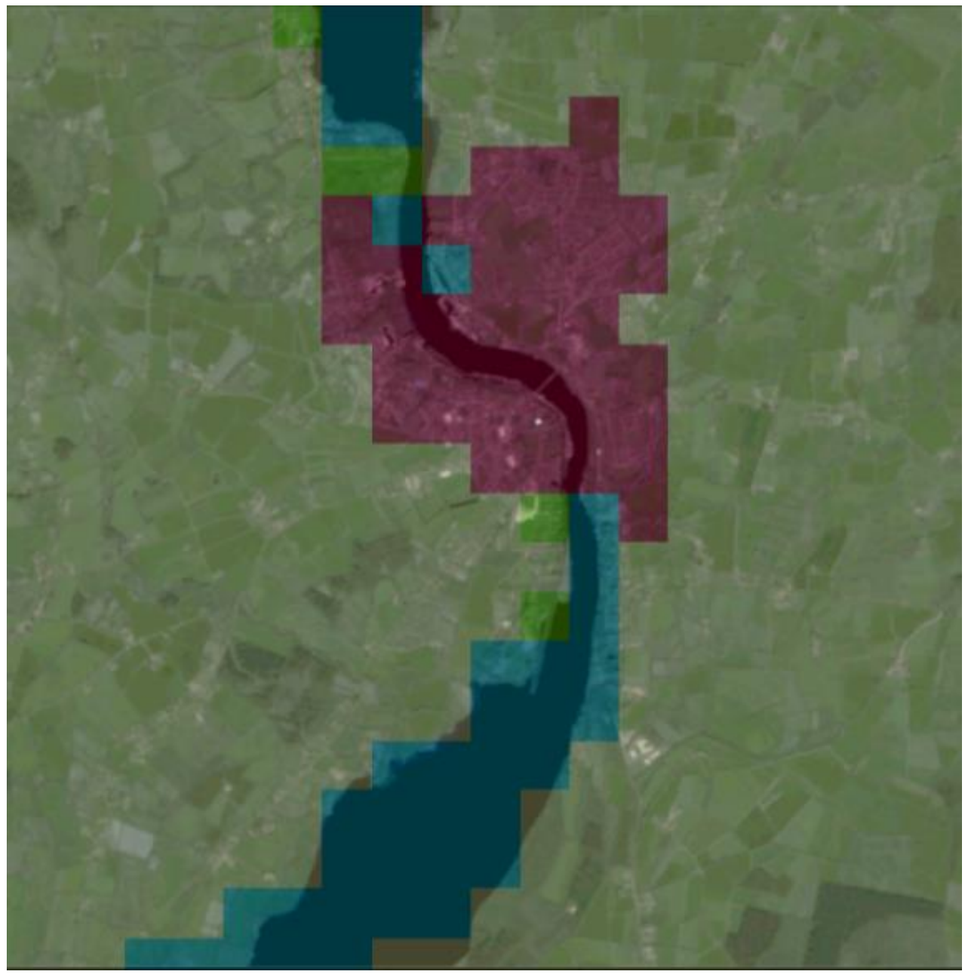


ECOCLIMAP-SG



Model Prediction

Primary Results - Resolution



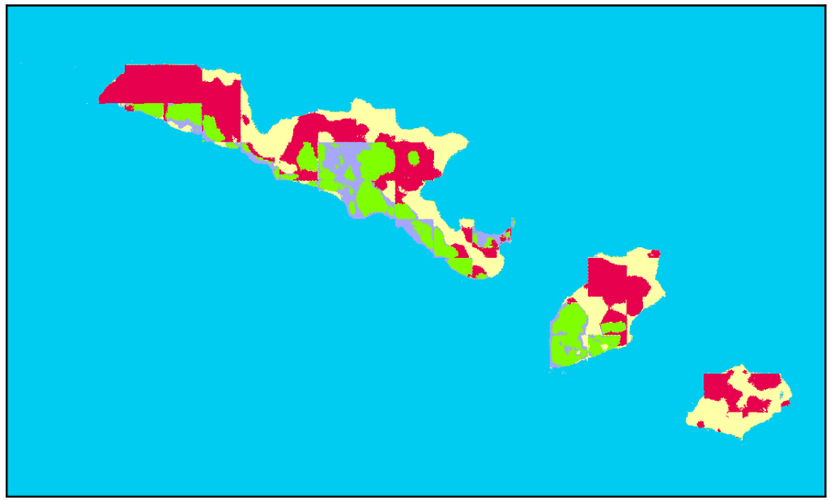
ECOCLIMAP-SG

Model Prediction

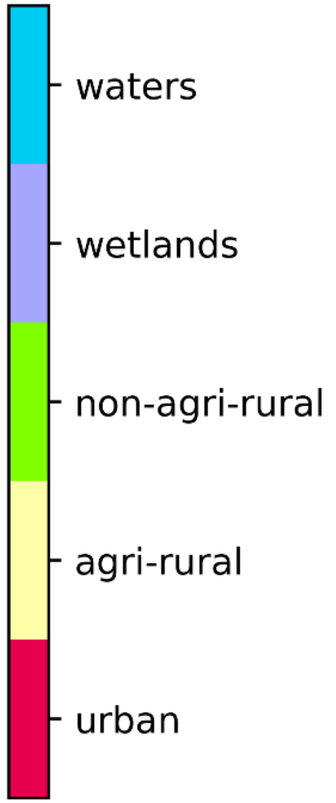
Interesting Observations

Aran Islands

MODEL PREDICTION



CORINE



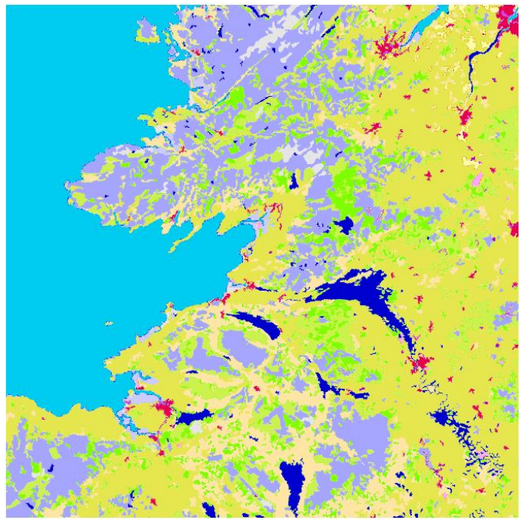
Secondary Training Data

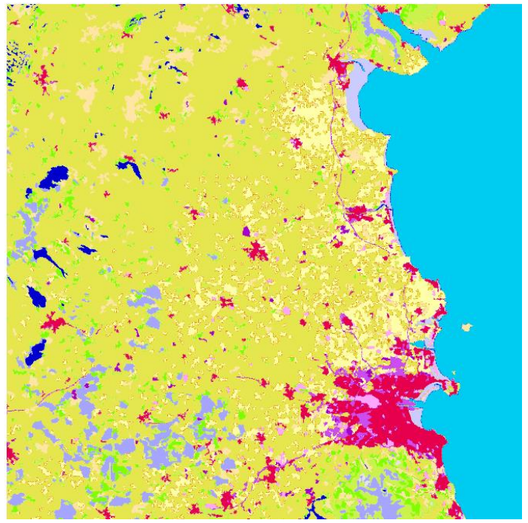
Donegal

Dublin

Clare

CORINE Segments

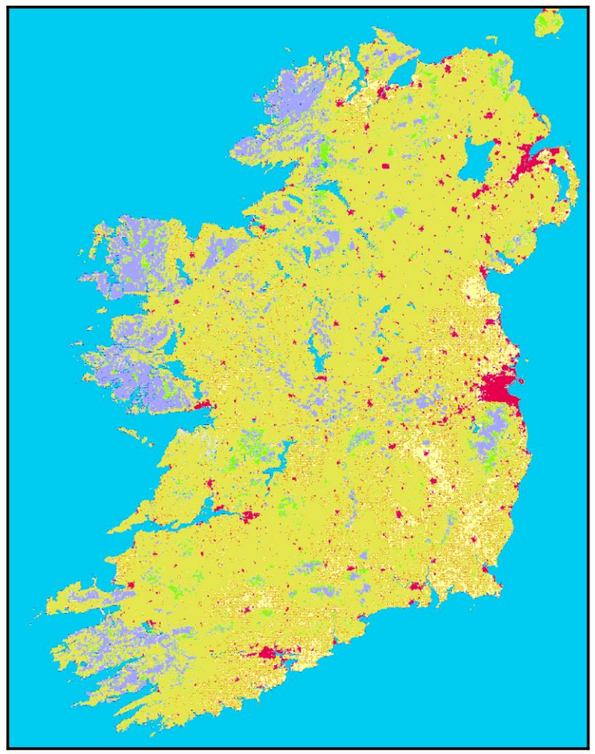




Sentinel-2 Segments

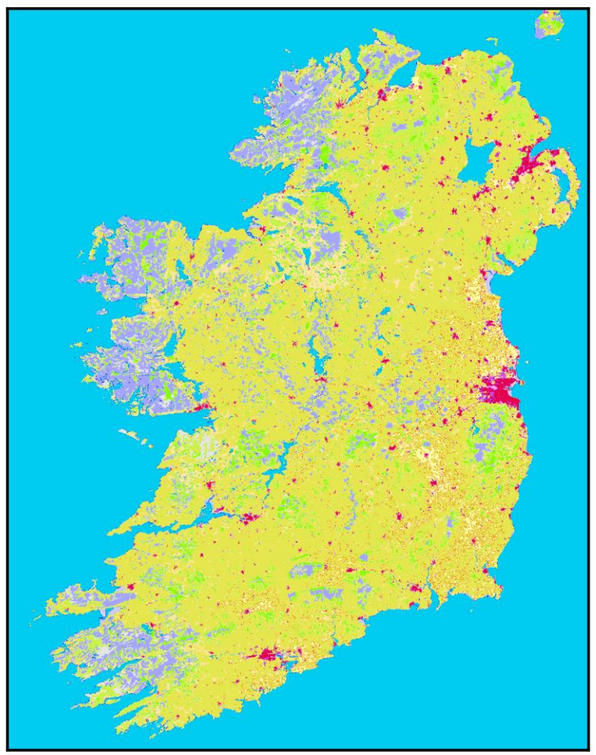
Secondary Results

ECOCLIMAP

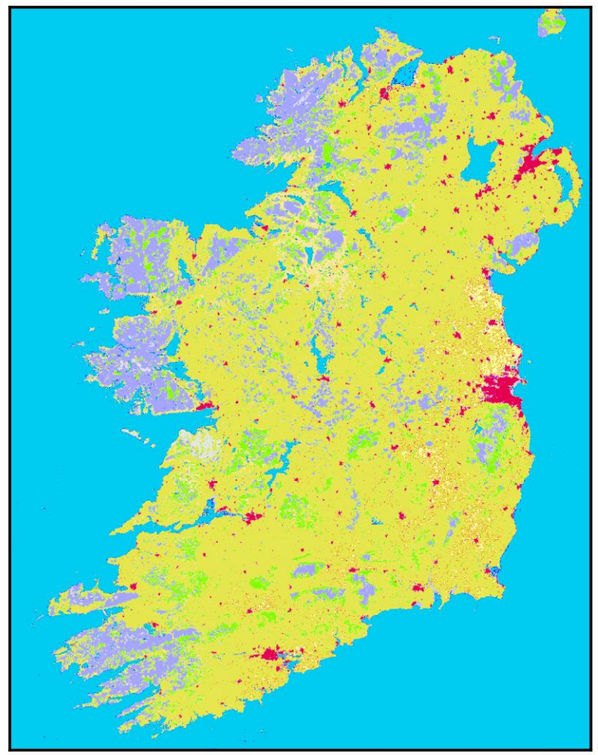


82.4%

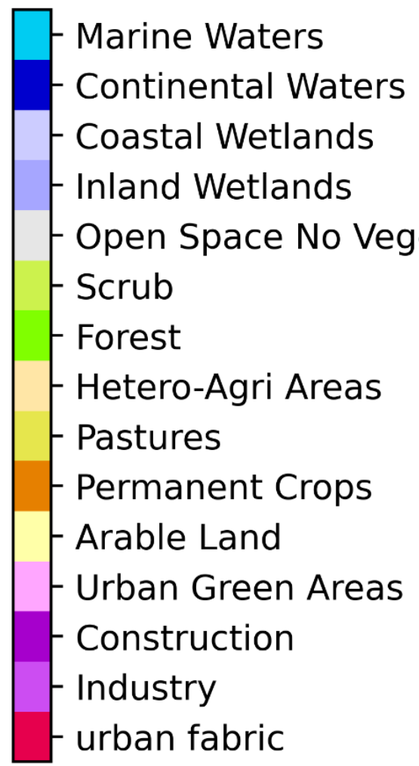
CORINE



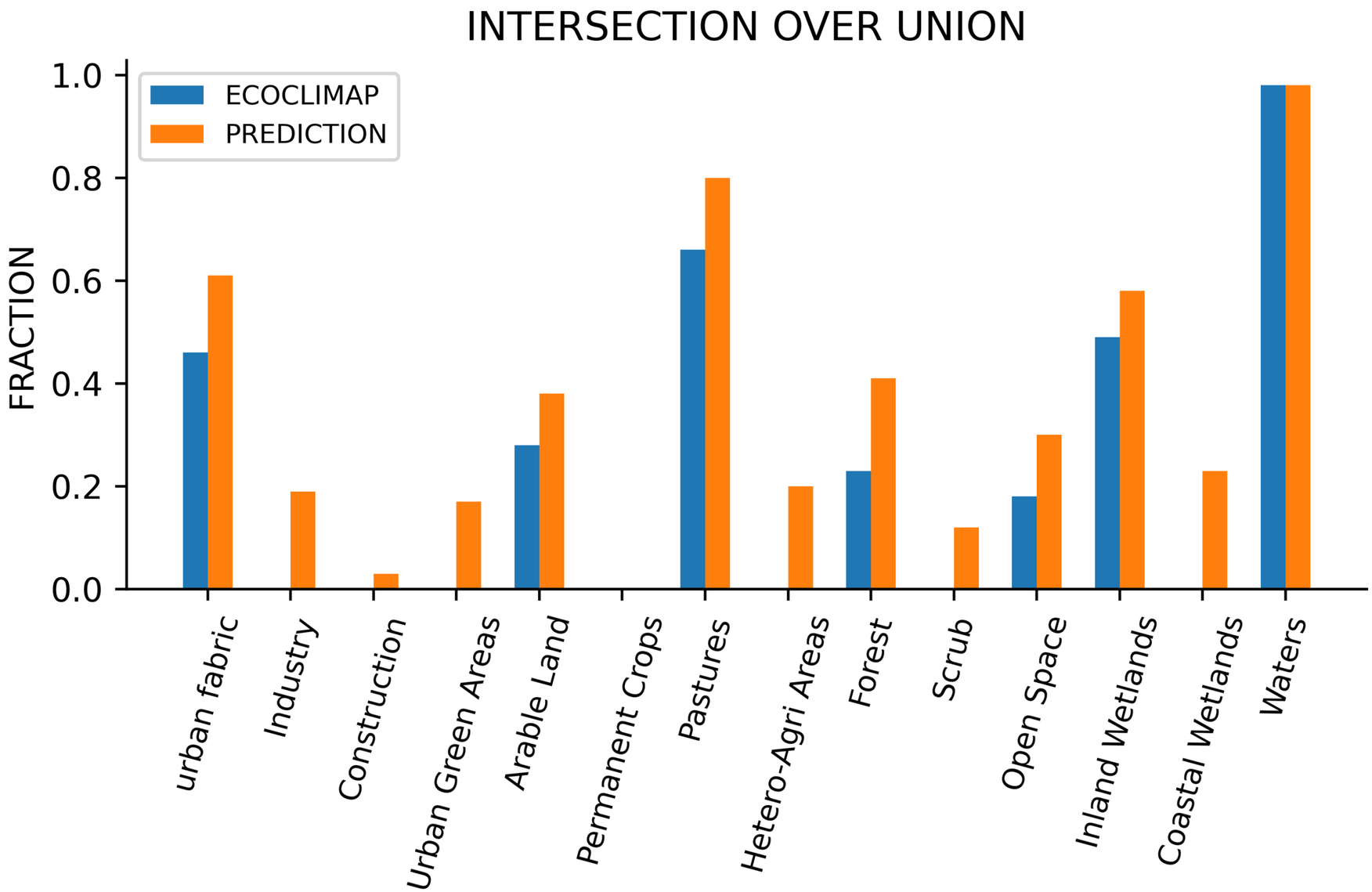
MODEL PREDICTION



87.3%



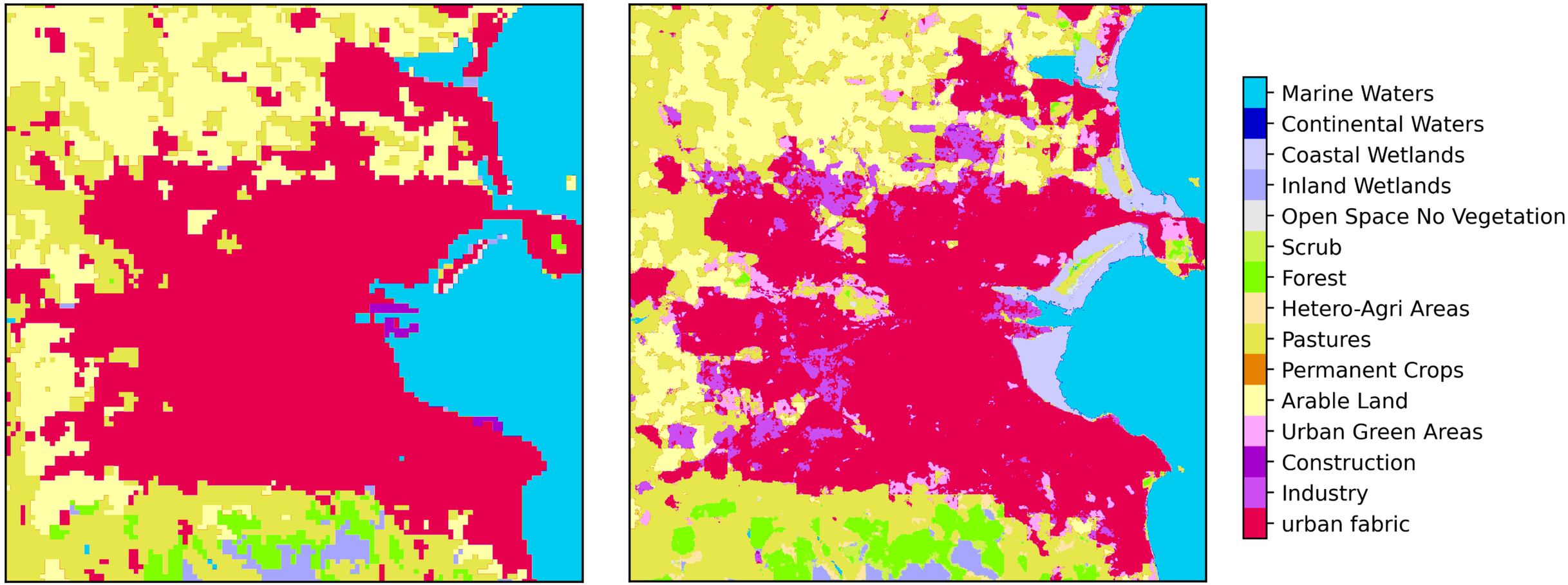
Secondary Results



Dublin

ECOCLIMAP

Model Prediction



Further Work

- Custom labels.
- More nuanced training data selection.
- Attempt a model trained across multiple jurisdictions.

Conclusions

- Machine Learning model shows an improved accuracy on ECOCLIMAP-SG.
- Machine Learning model offers a universal method of improving the map.
- Machine Learning shows a viable way of improving the resolution of the map.