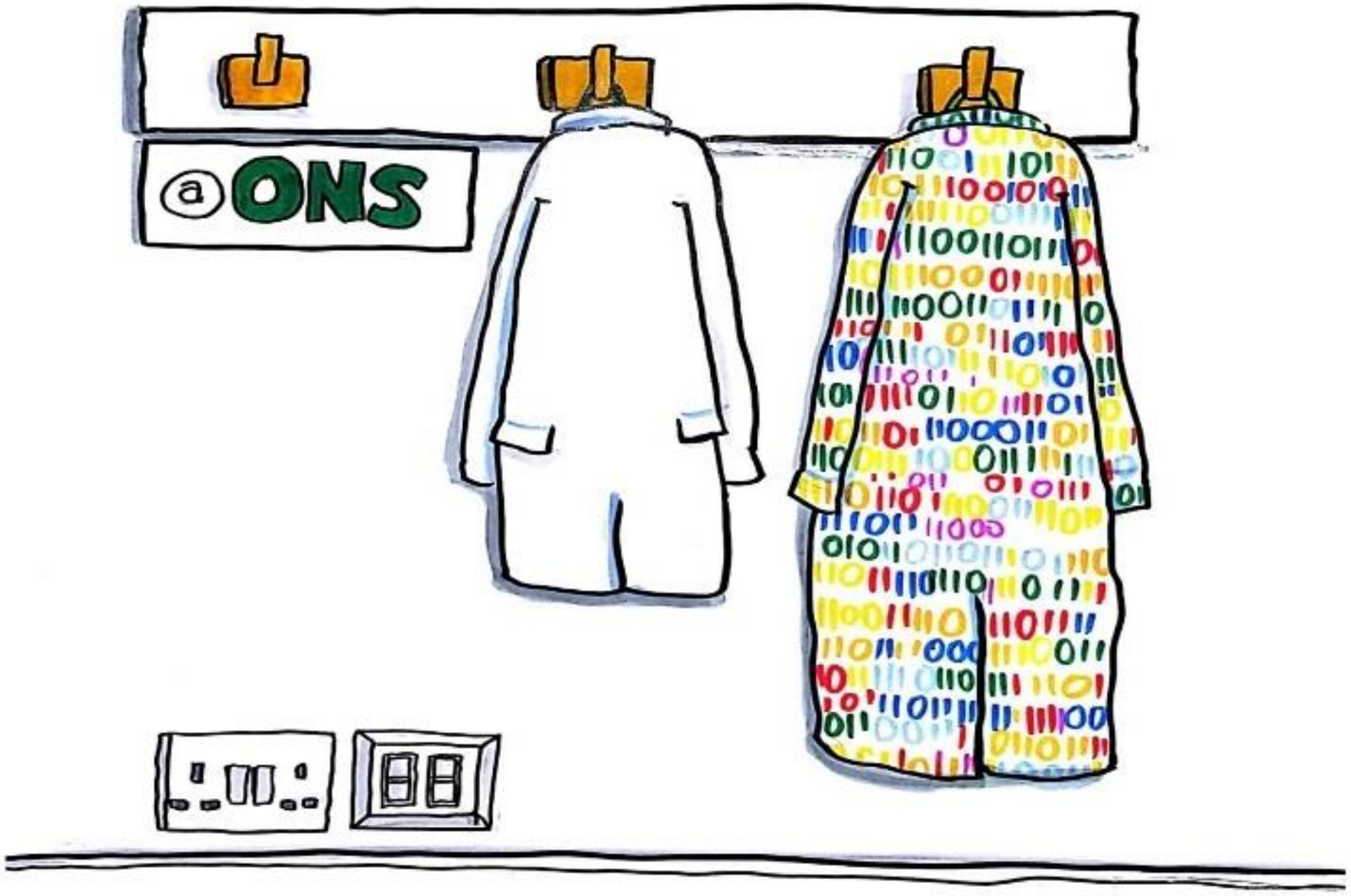
Green Spaces in Residential Gardens

Dr Christopher Bonham



Ordnance Survey

Data Science Campus

web: datasciencecampus.ons.gov.uk
email: datasciencecampus@ons.gov.uk
twitter: @DataSciCampus

The challenge

- Natural Capital. What % of UK gardens are covered in vegetation?

Aerial imagery

BlueSky - Public Sector Mapping Agreement (PSMA)

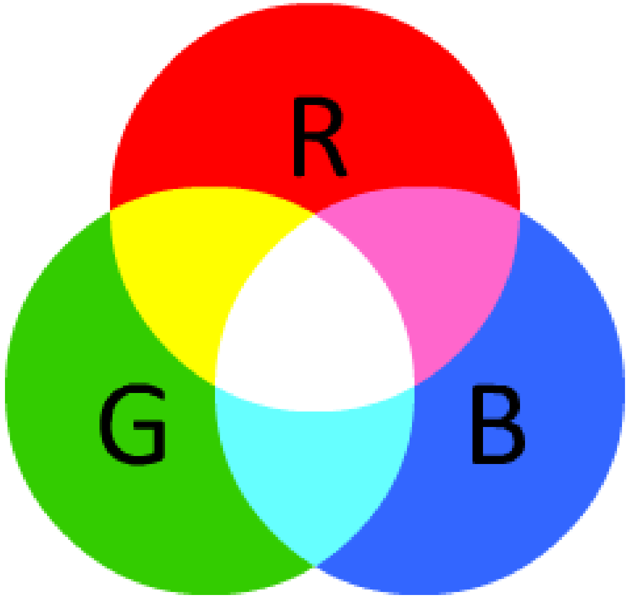
Private gardens polygons

Ordnance Survey MasterMap

Why?

- Flood risk mapping
  - Estimate the benefits of urban drainage such as vegetation.
- Urban heat calculation
  - Improving accuracy in model by including vegetation coverage
- House price predictions
  - Replace current estimation of green space
- Carbon footprint estimation
  - Differentiate between trees and grass

Off the shelf measures – RGB based



- All three metrics use the additive RGB colour system
- Equation is applied to the RGB values of each pixel
- Each metric ranges from 0 to 1
- Value > 0 - green living material

Visual Normalised Difference Vegetation Index (vNDVI)

- Uses only the green and red channels

$$vNDVI = \frac{(G - R)}{(G + R)}$$

Green Leaf Index (GLI)

- Development of vNDVI using the blue channel

$$GLI = \frac{(2 \times G - R - B)}{(2 \times G + R + B)}$$

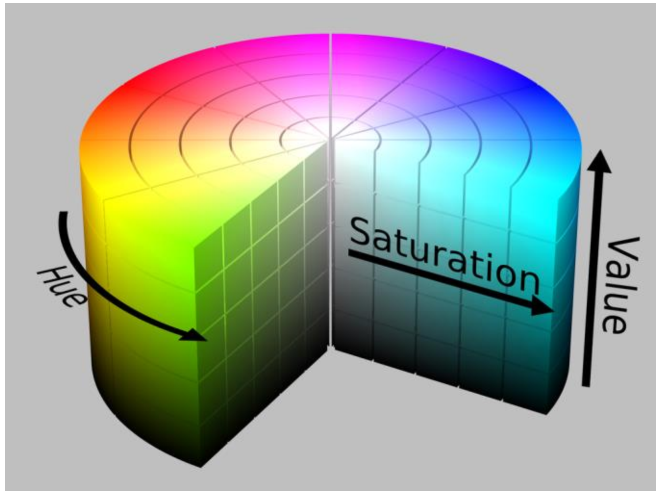
Visual Atmospheric Resistance Index (VARI)

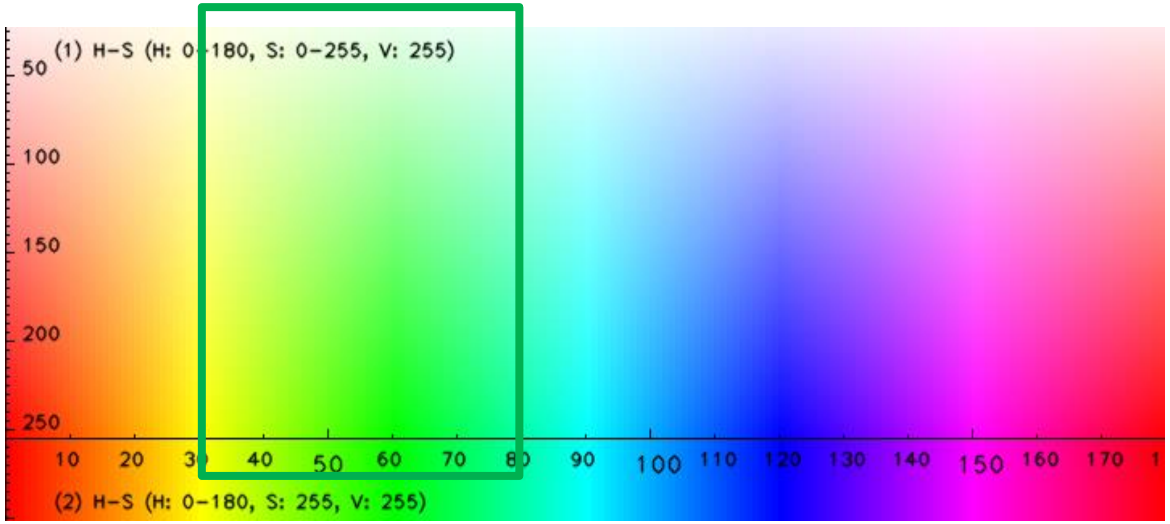
- Insensitive to atmospheric effects

$$VARI = \frac{(G - R)}{(G + R - B)}$$

Off the shelf measures – HSV based

- More closely aligned to human vision
- Hue (H), Saturation (S) and Value (V)
  - Hue represents colour
  - Saturation represents the amount of grey in the image
  - Value represents brightness
- Colour (hue) independent of saturation and value
- Pixel labelled as green if its hue is between 30 and 80

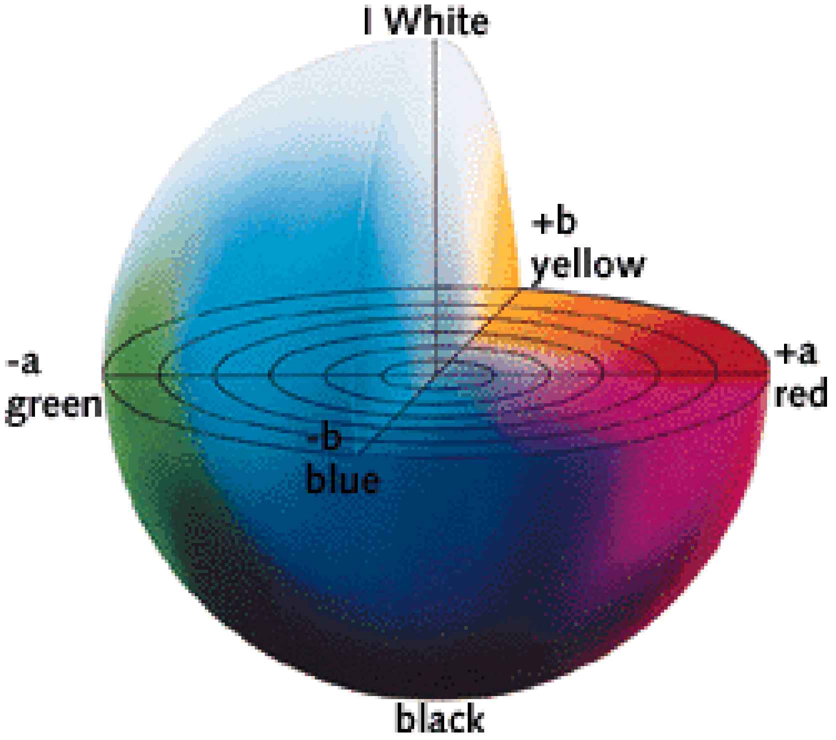




Off the shelf measures – Lab based

- CEILAB (Lab) colour space expresses colour as three numerical values
  - L represents lightness (black to white)
  - a represents green to red colours
  - b represents blue to yellow colours

- Previous work at the campus used labelled data from Mapilliary Vistas dataset to optimise a and b thresholds



Single Lab(a*) $-31 \leq a \leq -11$

Double Lab(a*b*) $-31 \leq a \leq -6$
$5 \leq b \leq 57$

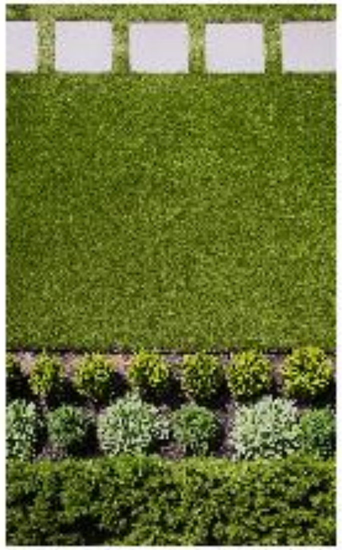
We don’t have ground truth

- Mapiliary Vistas image library classifies each pixel into a number of classes (people, car, tree, building etc)
- This can be used to assess performance of vegetation classification algorithms for street level imagery

- No equivalent labelled dataset for aerial imagery

- Initial study, library of 10 test images is used
- Each image represents a different garden feature
- Performance is quantitively assessed on these images

Results – Test image 1



- All approaches perform well
- Some noise in LAB results

vNDVI (84.2%)

GLI (88.4%)

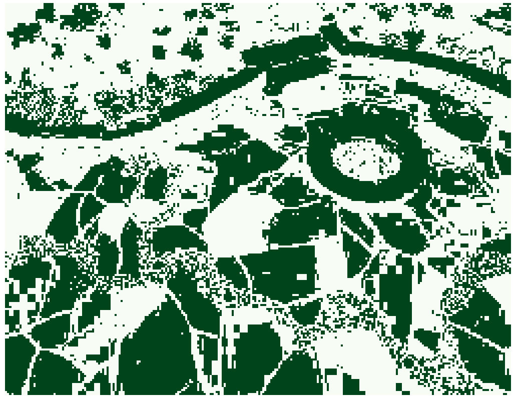
VARI (84.2%)

HSV (83.8%)

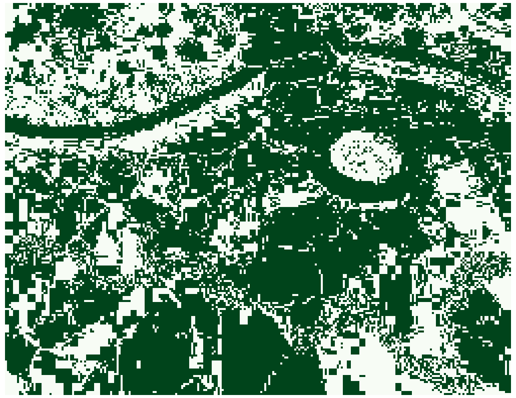
Lab(a*) (73.3%)

Lab(a*b*) (78.8%)

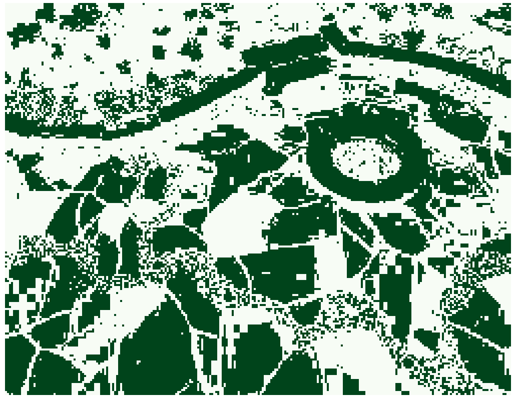
Results – Test image 3



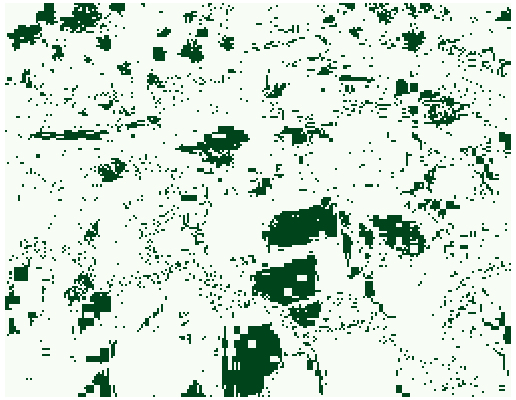
vNDVI (48.3%)



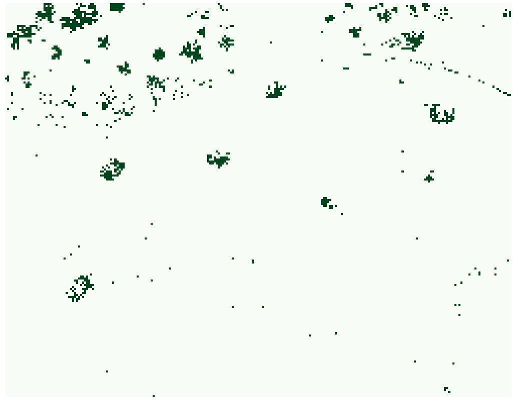
GLI (63.7%)



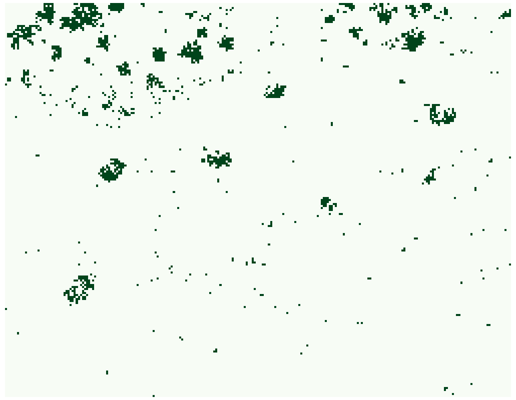
VARI (48.1%)



HSV (16.2%)



Lab(a*) (2.5%)

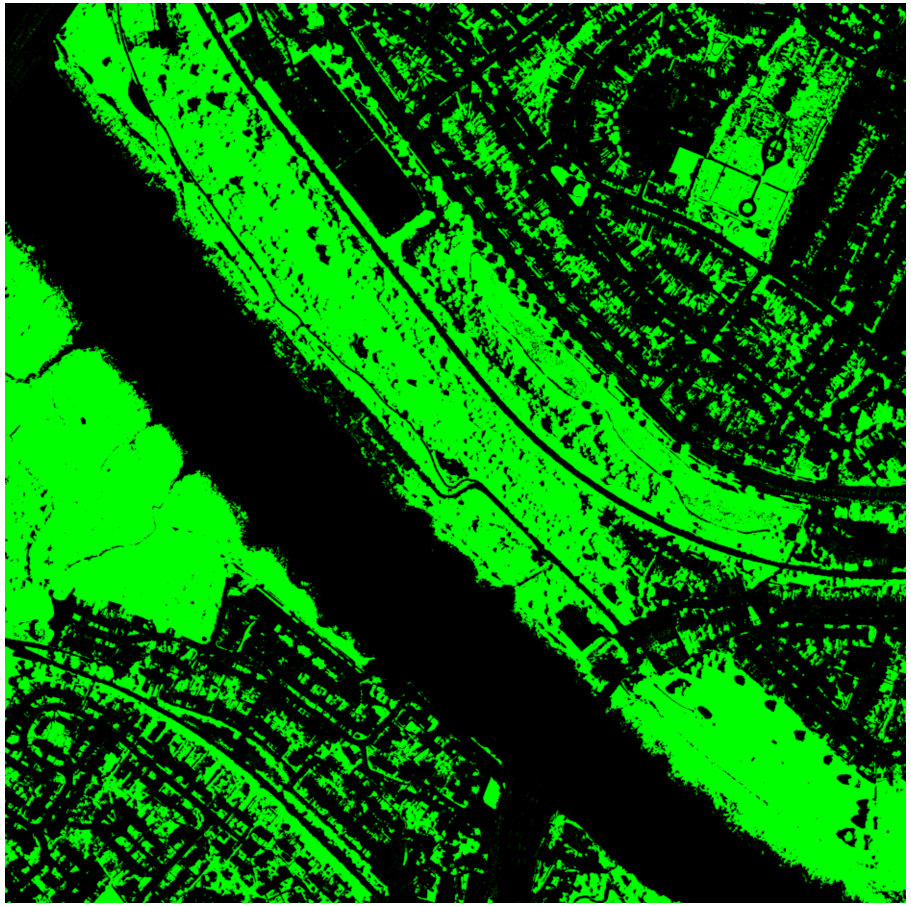


Lab(a*b*) (3.3%)

- Both LAB approaches perform well
- HSV classifies some slabs as green
- All RGB based algorithms very poor

Application of HSV to Bristol – High Level

Original Picture (25cm)



Green pixels (25cm)
HSV color scheme
30 ≤ Hue threshold ≤ 80

Application of HSV to Bristol – Low level

Original Picture (12.5cm)

Garden identified by polygon

Green pixels
HSV color scheme
30 ≤ Hue threshold ≤ 80

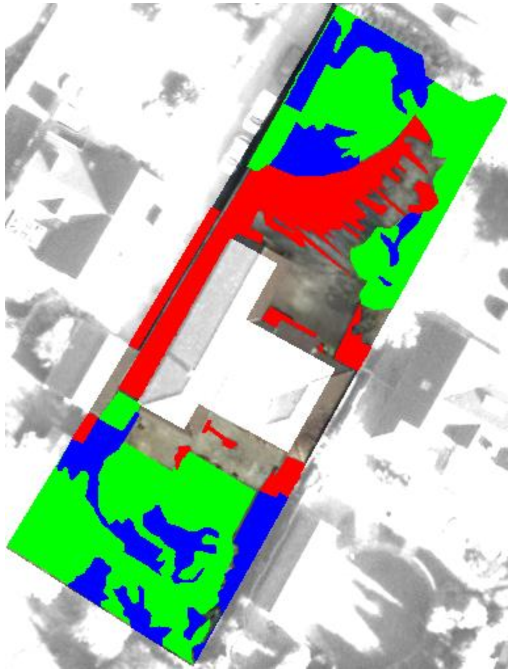
Supervised learning

We need labels to train algorithm

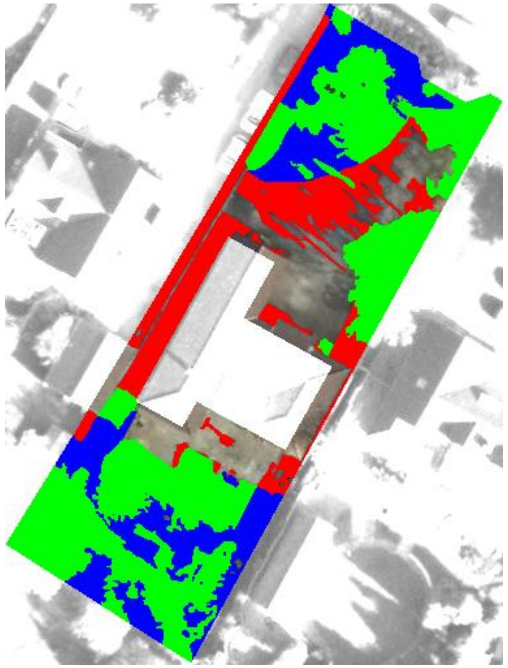
- 100 images randomly selected from Cardiff and Bristol
- 4 people each manually labelled 75 images (100 images labelled by 3 different people)
  - **Green** – vegetation
  - **Blue** – vegetation in shade
  - **Red** – urban in shade
- Voting logic classifies each pixel based upon the most popular choice, out voting bad choices
- Classification matrix gives agreement rate between reviewers
- If agreement rate drops below 80% fourth person reviews image

Manual labelling

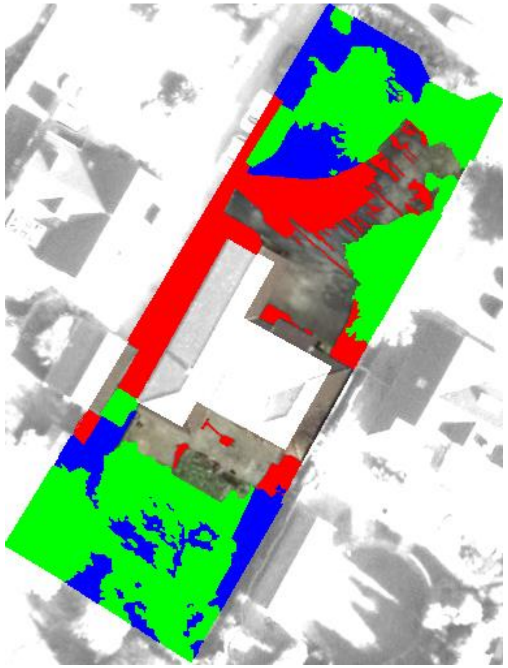
Original Image



Review 1



Review 2



Review 3

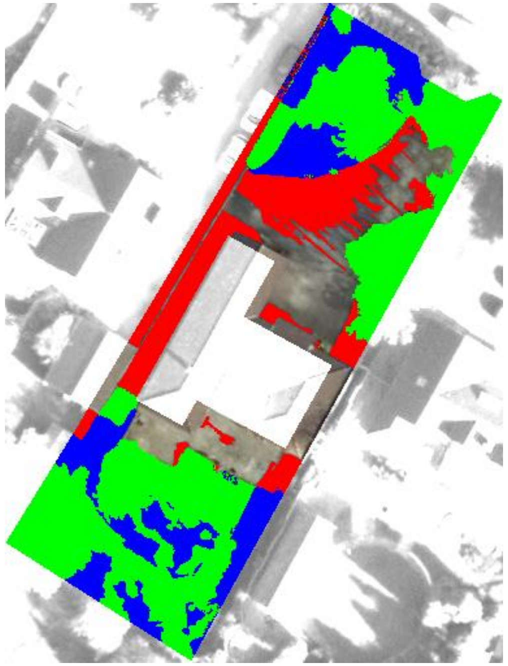
vegetation | vegetation in shade | urban in shade

Manual labelling

Original Image

	1	3	4
1		85%	84%
3	85%		87%
4	84%	87%	

Agreement matrix



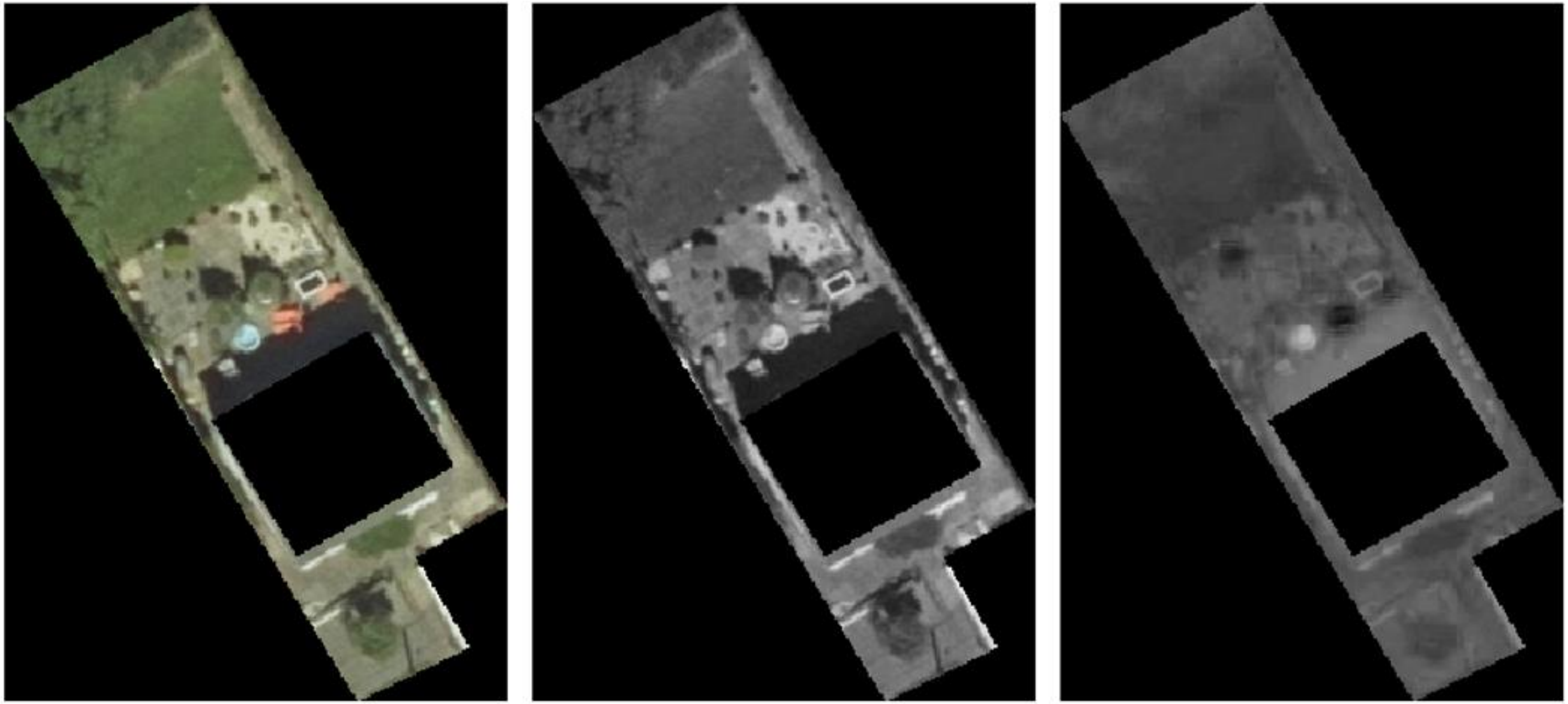
Resulting labels

vegetation | vegetation in shade | urban in shade

Neural network architecture

- Three layer 12:8:4 network
- Standard ReLU activation units
- Output softmax activation units
- 12 features
  - Red, green, blue and infra-red spectral channels
  - Monochromatic principal component with first component removed
  - Three brightness principal components with first component removed
  - Three colour principal components with first component removed
- 4 Outputs, probability pixel is
  - Vegetation
  - Vegetation in shade
  - Urban in shade
  - None of the above

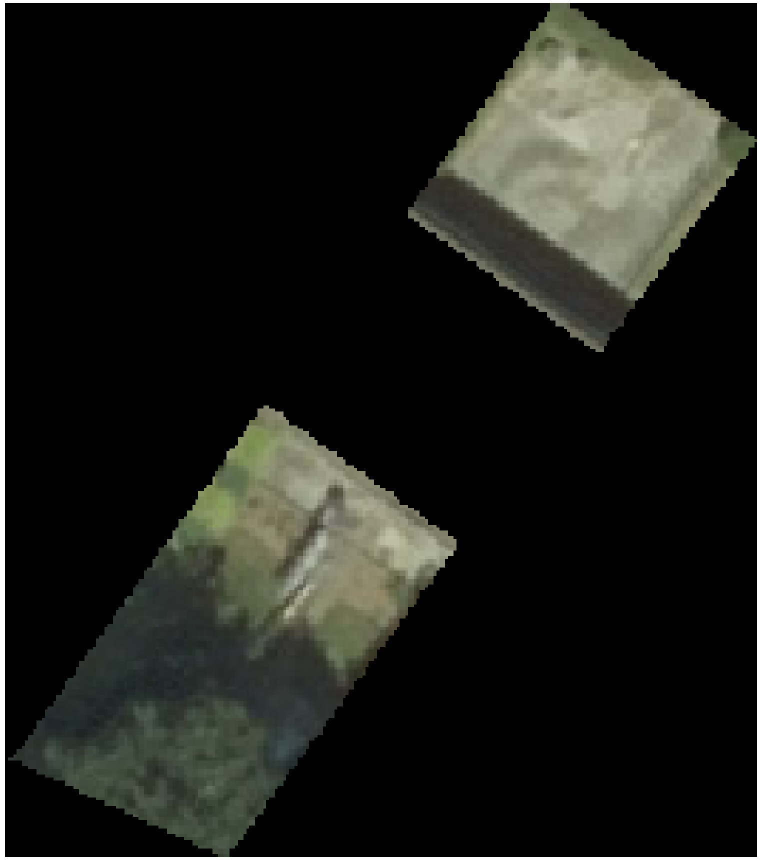
Isolate and remove shadows



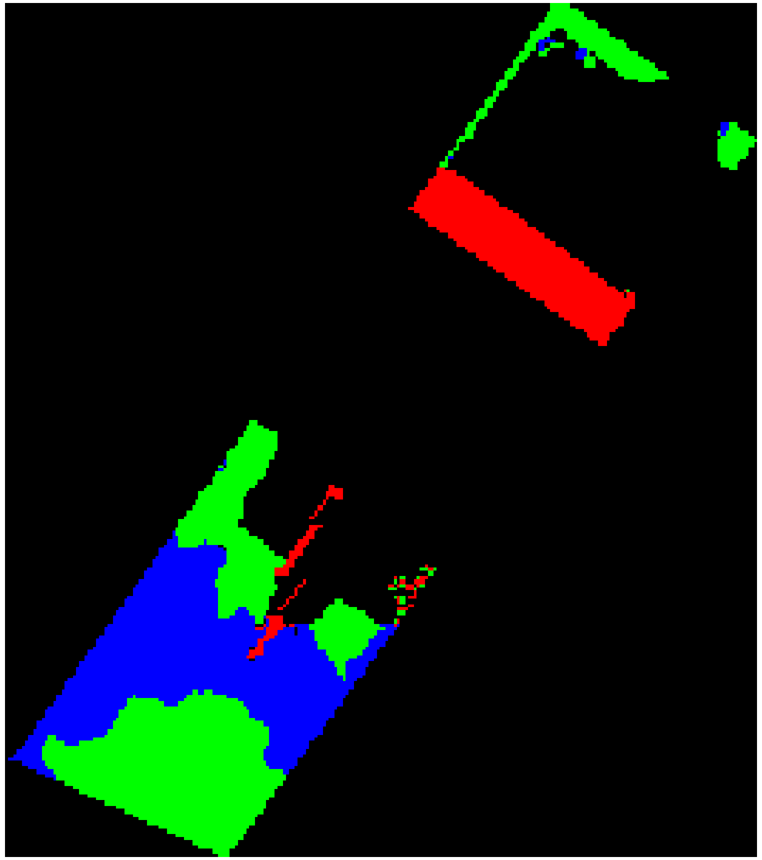
{R, G, B} | {R, G, B} → PC1 | {R, G, B} → {PC1, PC2}

Neural network classifier - results

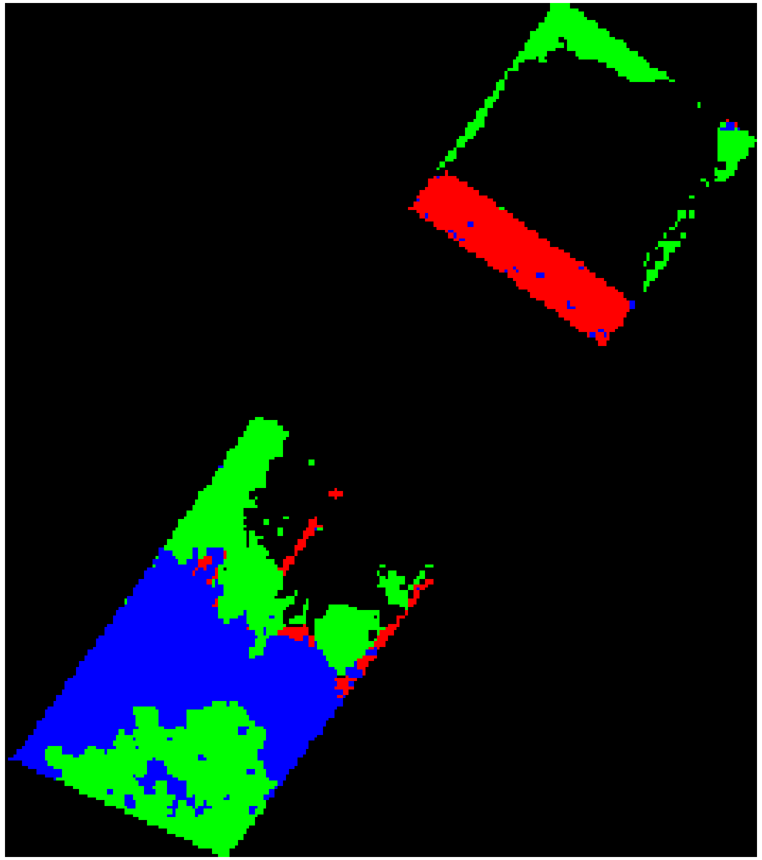
Original Image



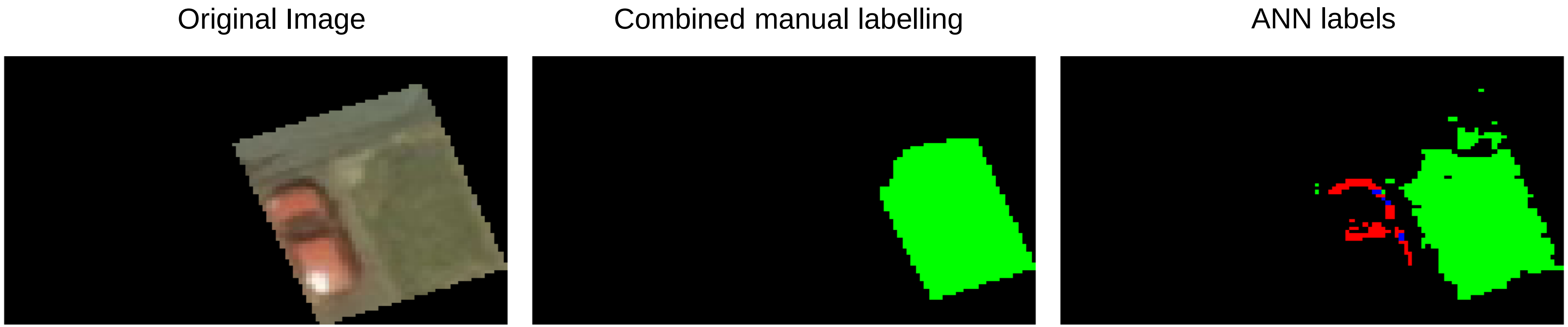
Manual labelling



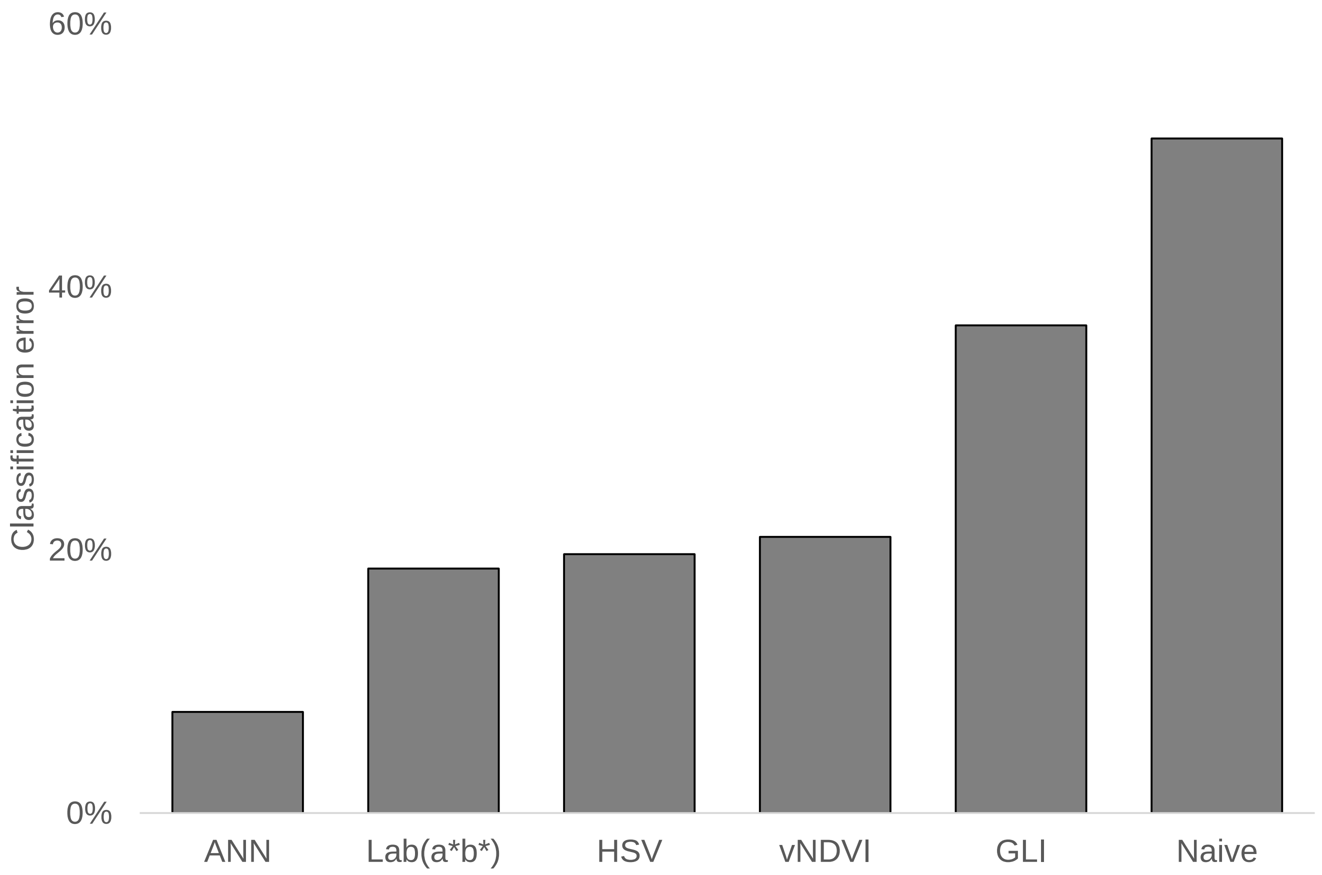
ANN predicted labels



Classifier corrected our mistakes!



Neural network classifier - results



Classification error across labelled images

Sample results

Cardiff

$13.4 km^2$ garden area
53.9% vegetation

Bristol

$41.9 km^2$ garden area
45.0% vegetation

Deployment and next steps

- Model deployed within a Python pipeline
- Run time for entire UK just over a day
  - Two-core virtual machine hosted on a Xeon E5-2650 @ 2.20GHz; 4Gb of memory
- Code available on GitHub

Potential developments

- Look at temporal trends
- Don't treat pixels in isolation (consider local neighbours)
- Account for weather conditions and seasonality
- Expand classification to identify specific types of organic material
- Deployment on the cloud