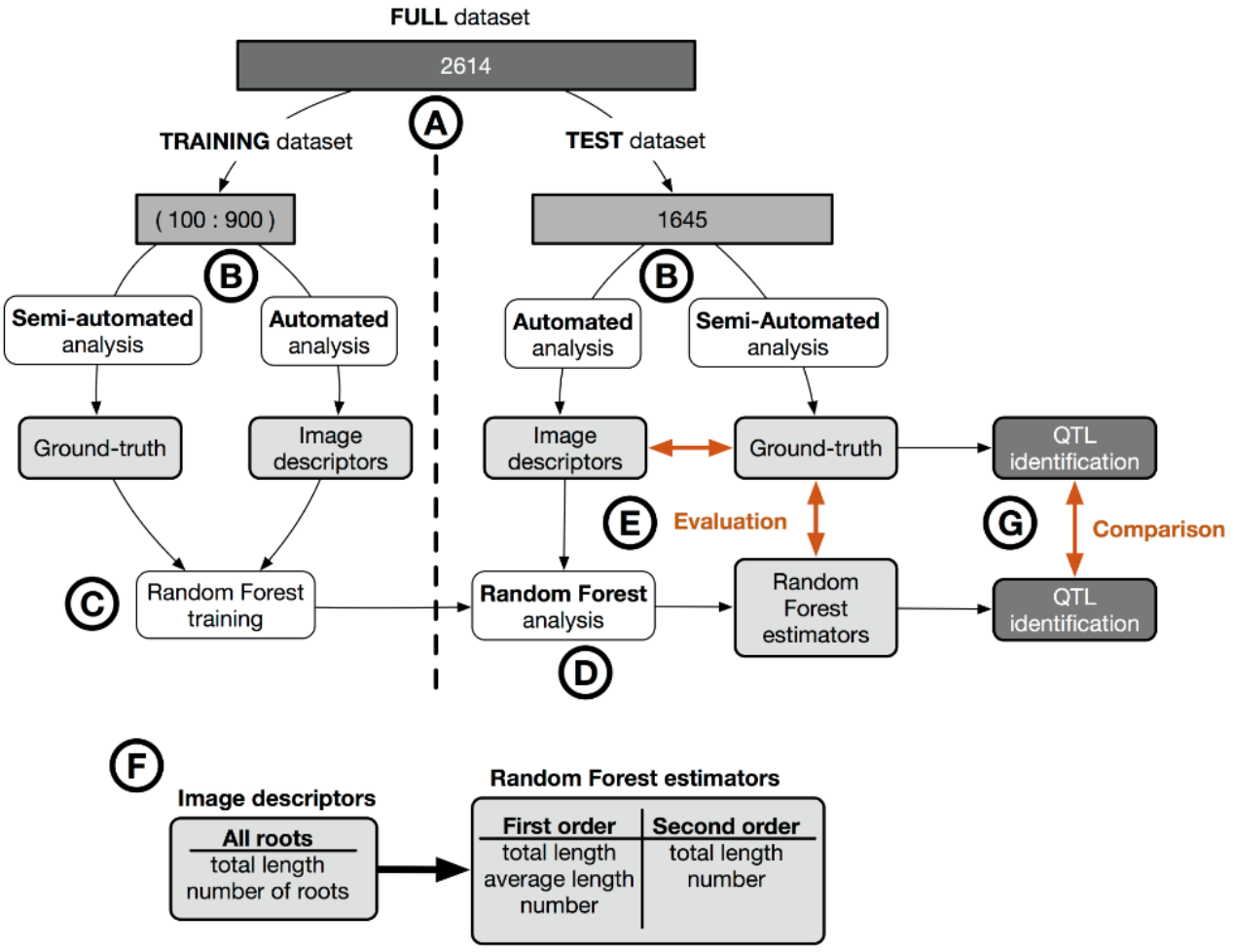
PRIMAL: Pipeline of Root Image analysis using MAchine Learning

Guillaume Lobet, Jonathan A. Atkinson, Manuel Noll, Markus Griffiths, Darren M. Wells

Abstract



Citation: Guillaume Lobet, Jonathan A. Atkinson, Manuel Noll, Markus Griffiths, Darren M. Wells PRIMAL: Pipeline of Root Image analysis using MAchine Learning. **protocols.io**

https://www.protocols.io/view/primal-pipeline-of-root-image-analysis-using-machi-h7bb9in

Published: 19 Jun 2017

Protocol

Extract image descriptors

Step 1.

Use a fully automated tool to extract image descriptors global dataset.

Automated tools that we recommend:

- RIAJ (https://github.com/guillaumelobet/Root-Image-Analysis-Pipeline-Evaluation/tree/master/01_image_analysis)

- DIRT (http://dirt.iplantcollaborative.org/welcome)

Here is an example of a descriptor file:

https://github.com/plantmodelling/primal/blob/master/www/global_estimators.csv

EXPECTED RESULTS

Data table containing image descriptors for the all image library

The datatable should contain an unique identifier for each image.

id	descriptor 1	....	descriptor i
image 1			
...			
image j			

Create the GROUND-TRUTH SET

Step 2.

Use a semi-automated tool to extract the ground-truth out of 200 random images (the **GROUND-TRUTH SET**).

Semi-automated tools that we recommend:

- RootNav (https://www.cpib.ac.uk/tools-resources/software/rootnav/)

- SmartRoot (https://smartroot.github.io/)

- EZ-Rhizo (http://www.psrg.org.uk/plant-biometrics.html)

- Root System Analyser (https://www.csc.univie.ac.at/rsa/)

Here is an example of a ground-truth file:
https://github.com/plantmodelling/primal/blob/master/www/global_estimators.csv

EXPECTED RESULTS

The datatable contains ground-truth data for the subset of the image library.

Each row should have an unique identifier (id) that correspond to the ones in the descriptor datatable

Launch the PRIMAL app

Step 3.

- Open an R console (for instance using RStudio)
- Type the following commands:

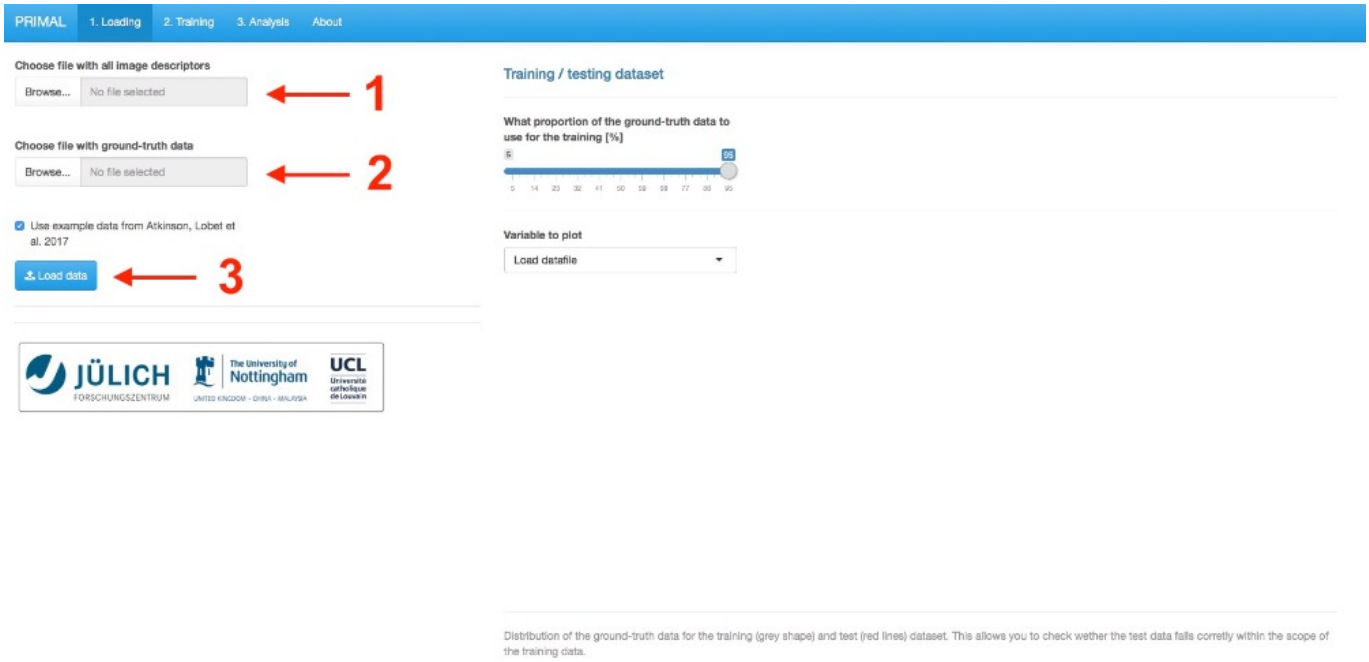
```
library(shiny)
shiny::runGitHub('plantmodelling/primal', 'plantmodelling')
```

This will launch the app in your browser (it still runs locally).

Load your data

Step 4.

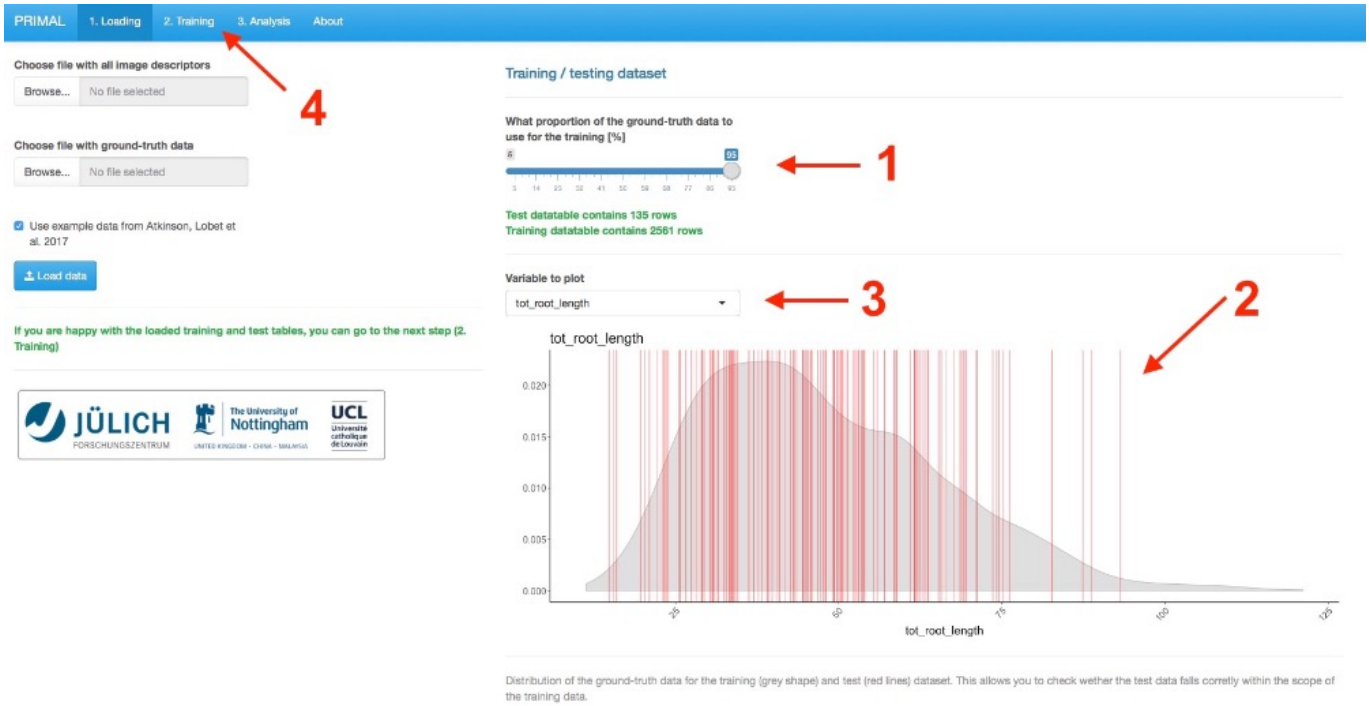
1. Load the DESCRIPTOR dataset
2. Load the GROUND-TRUTH dataset
3. Click Load



Choose the training data

Step 5.

1. Select what proportion of the GROUND-TRUTH dataset you want to use for the training (by default 95%). The rest will be used for the testing.
2. Look if the testing data (red lines) fall within the range of the training data (grey area)
3. Change the type of data to visualize
4. Go to the next step

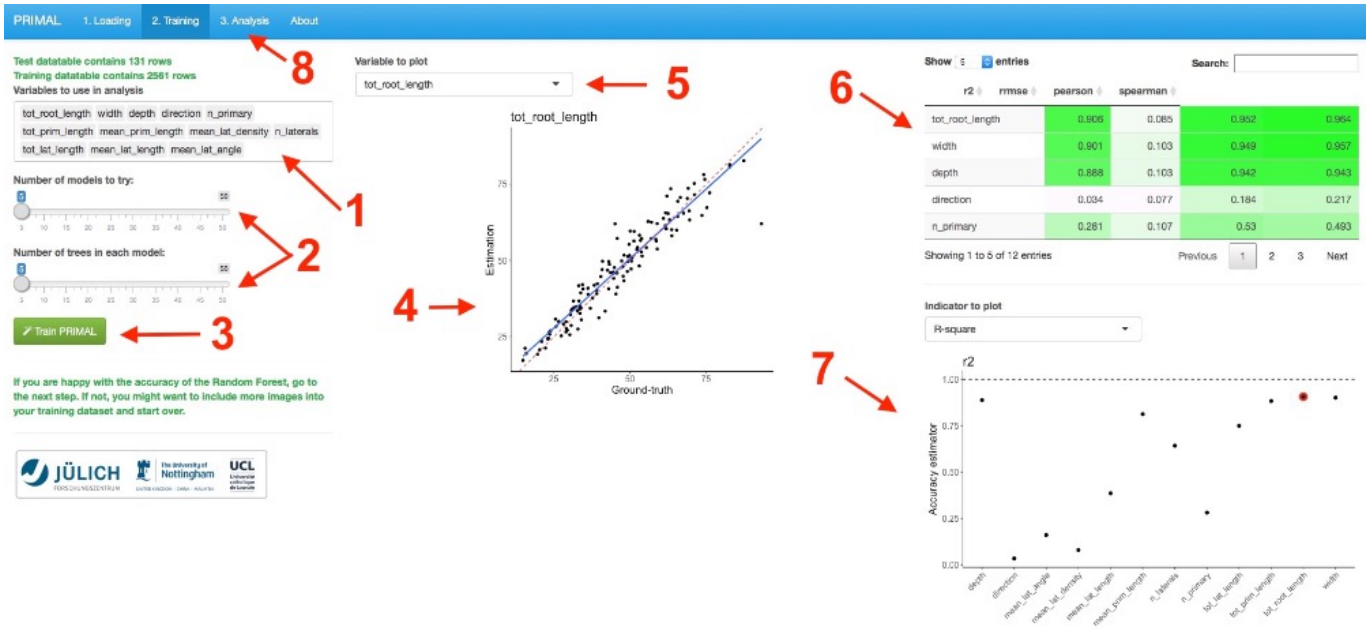


Train you data

Step 6.

1. Select the type of data you would like to evaluate with the Random Forest algorithm
2. Select the parameters for the Random Forest. The bigger the number of trees and models, the better the prediction. But also the slower.
3. Run the training (this might take a while. Take a coffee)
4. Look at the Random Forest estimation accuracy (vs the ground-truth). This is done using the test dataset
5. Change the parameter to look at
6. Table with different accuracy metrics for the different parameters
7. Visual representation of the accuracy metrics
8. If you are happy with the accuracy of the prediction, go to next step

IF THE ACCURACY IS NOT GOOD ENOUGH, GO BACK TO STEP 2 AND INCREASE THE SIZE OF THE **GROUND-TRUTH** DATASET



Apply the model to whole dataset

Step 7.

1. Click "Unleash PRIMAL"
2. Download the machine learning estimators.