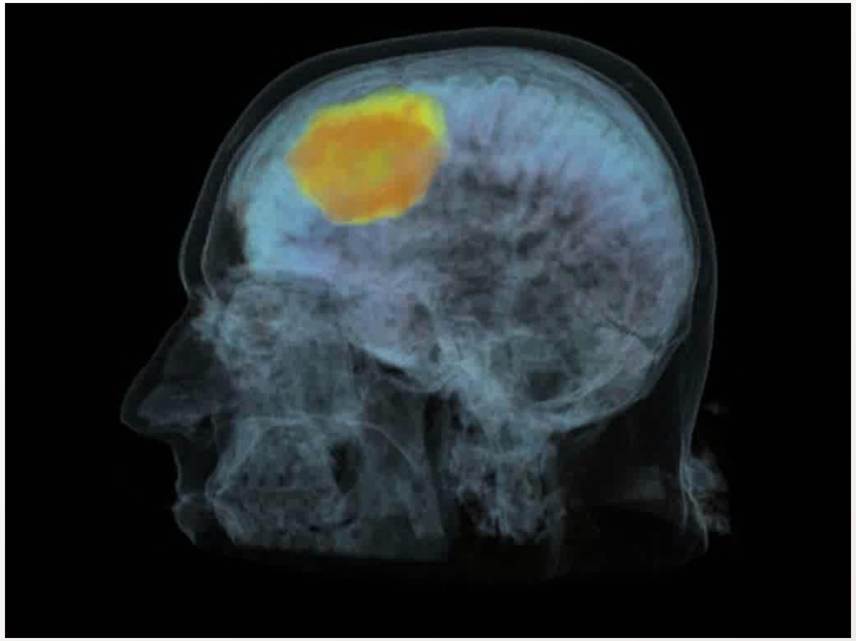
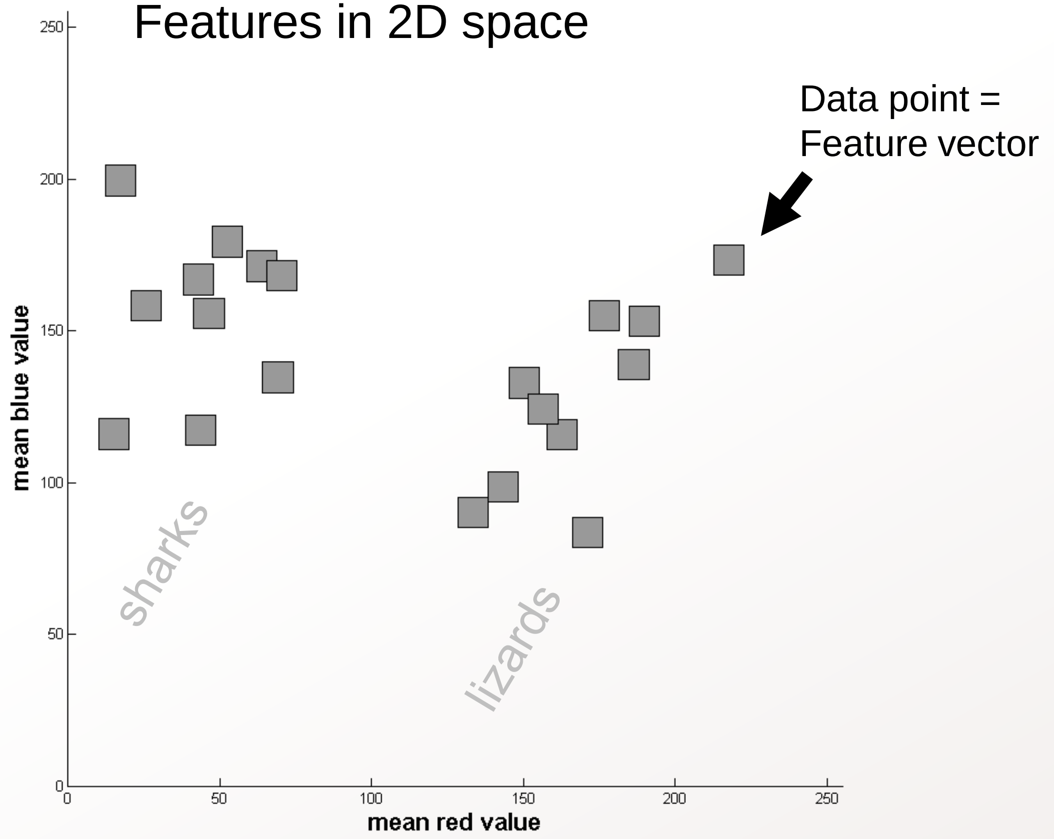
Machine Learning in Health Care



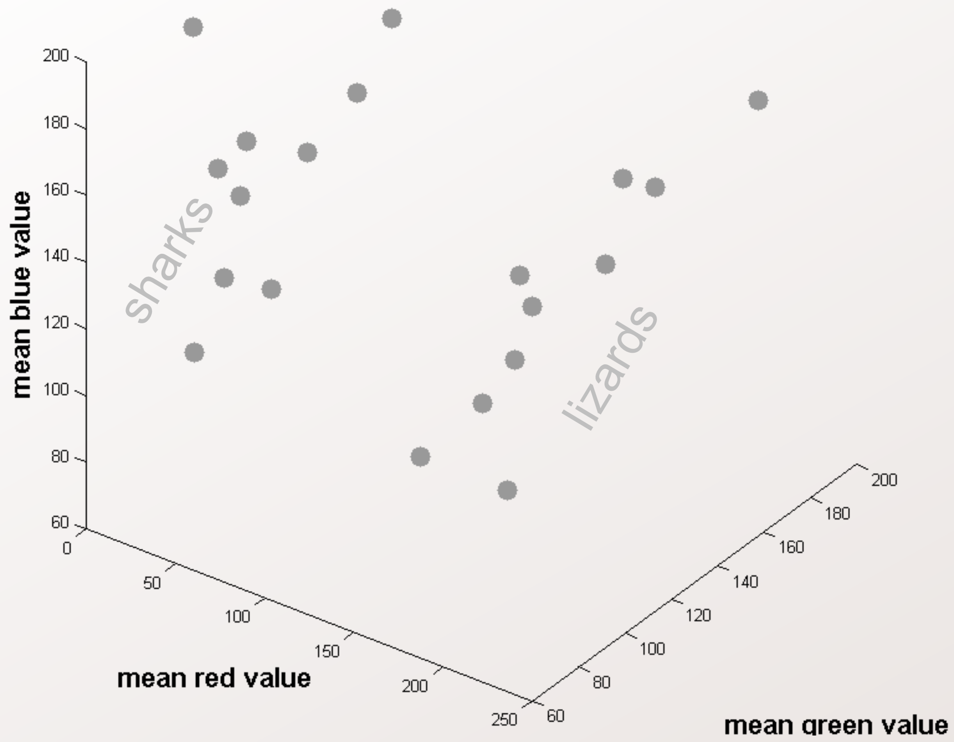
A. Criminisi

Machine learning for image content recognition

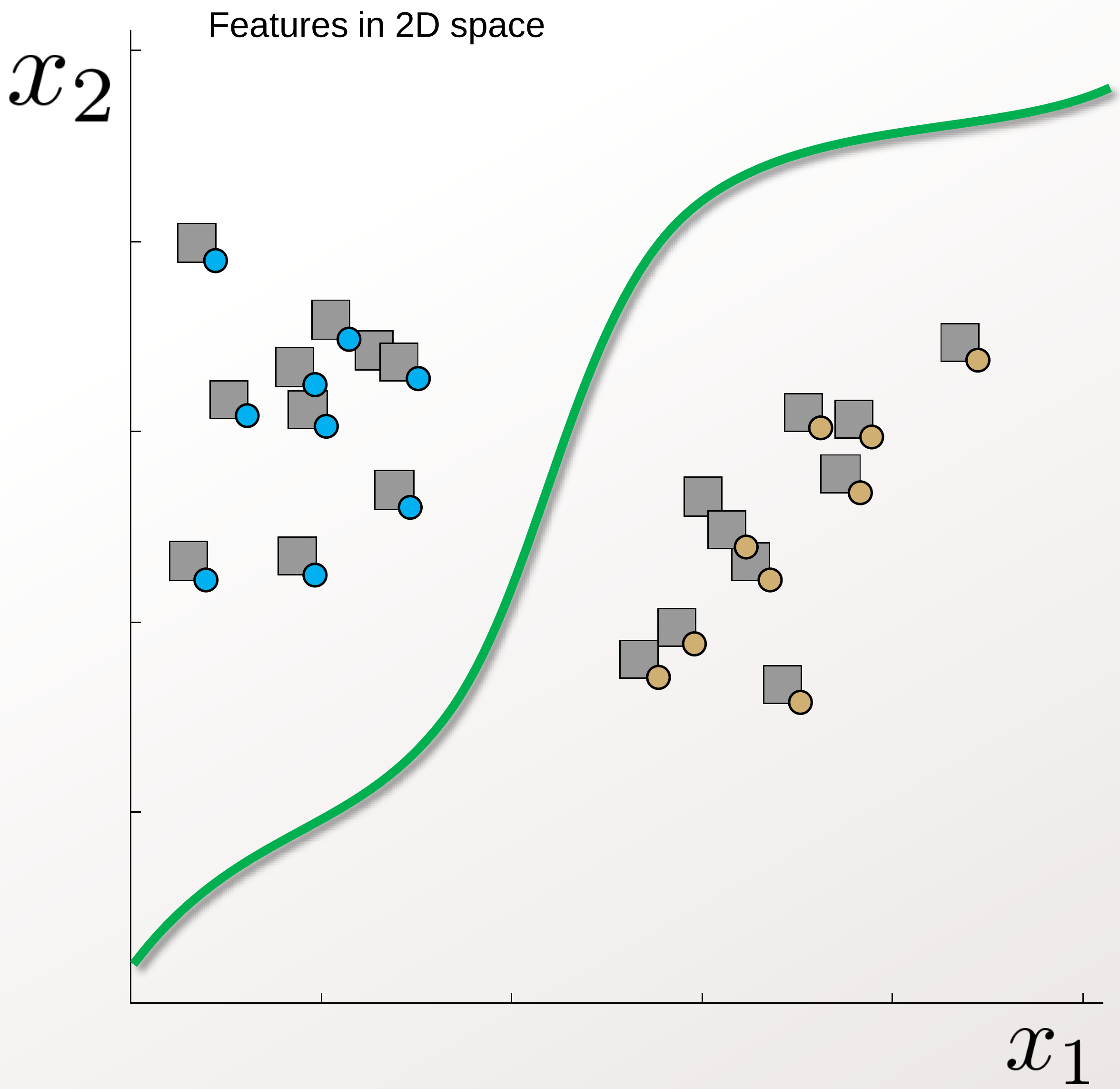
Features in 2D space



Features in 3D space



Machine learning for image content recognition



induction

Application: Kinect for Xbox gaming

Task: assigning body part labels to each pixel in Kinect-acquired depth videos

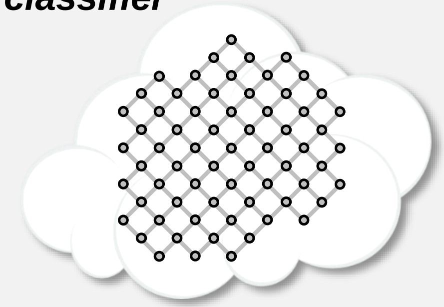
Input test depth image

Body part segmentation

image measurements made relative to pixel

e.g. depth, color, neighbors

classifier

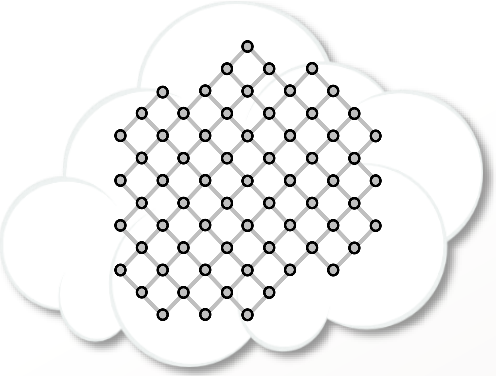


per-pixel prediction of class label

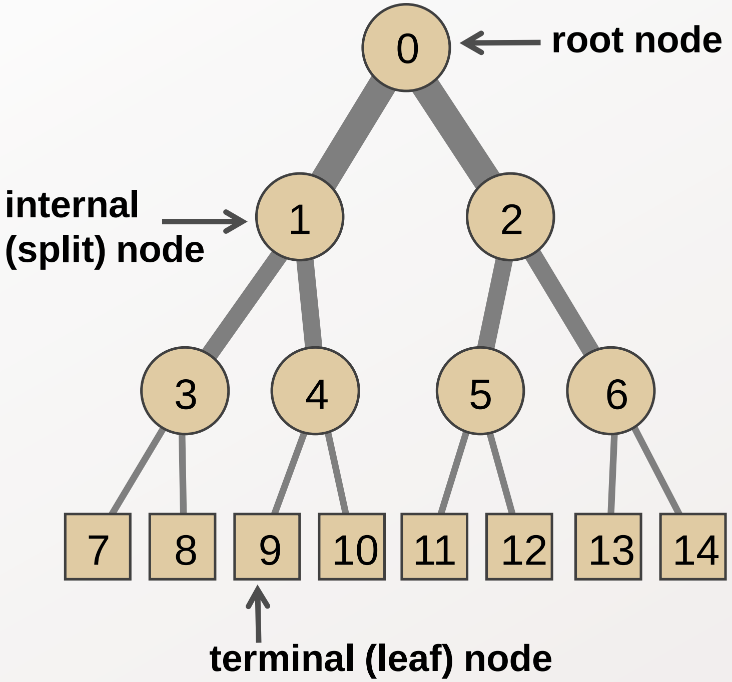
J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake, **Efficient Human Pose Estimation from Single Depth Images**, in *Trans. PAMI*, IEEE, 2012

Decision trees

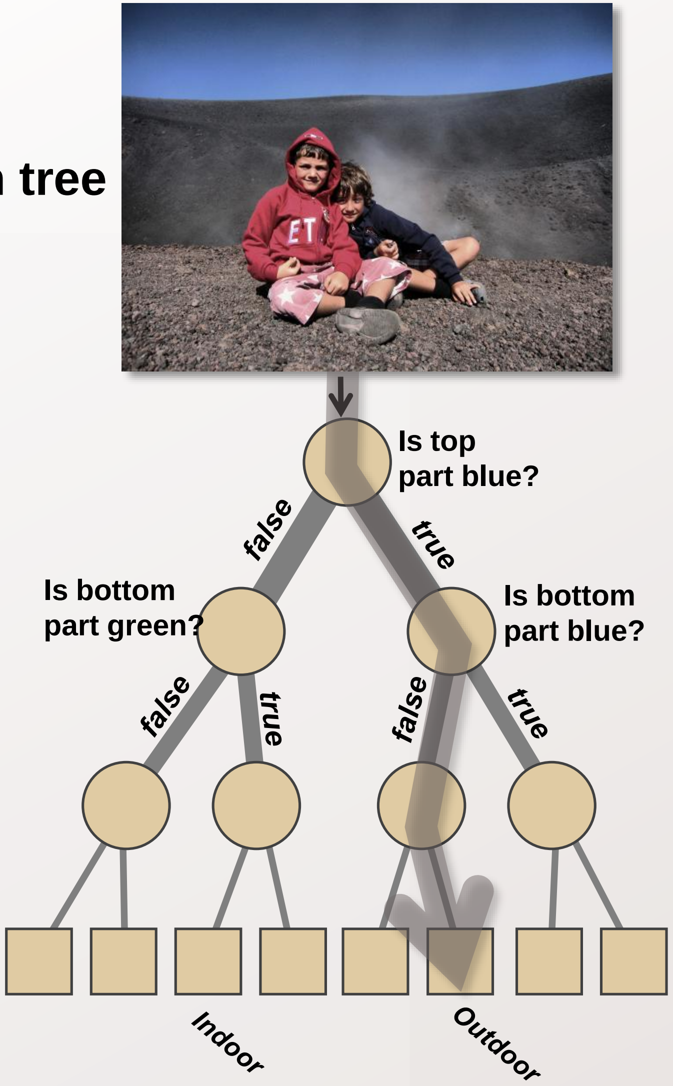
A general learned predictor



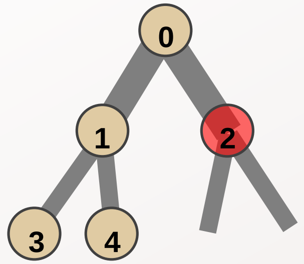
A general (binary) tree structure



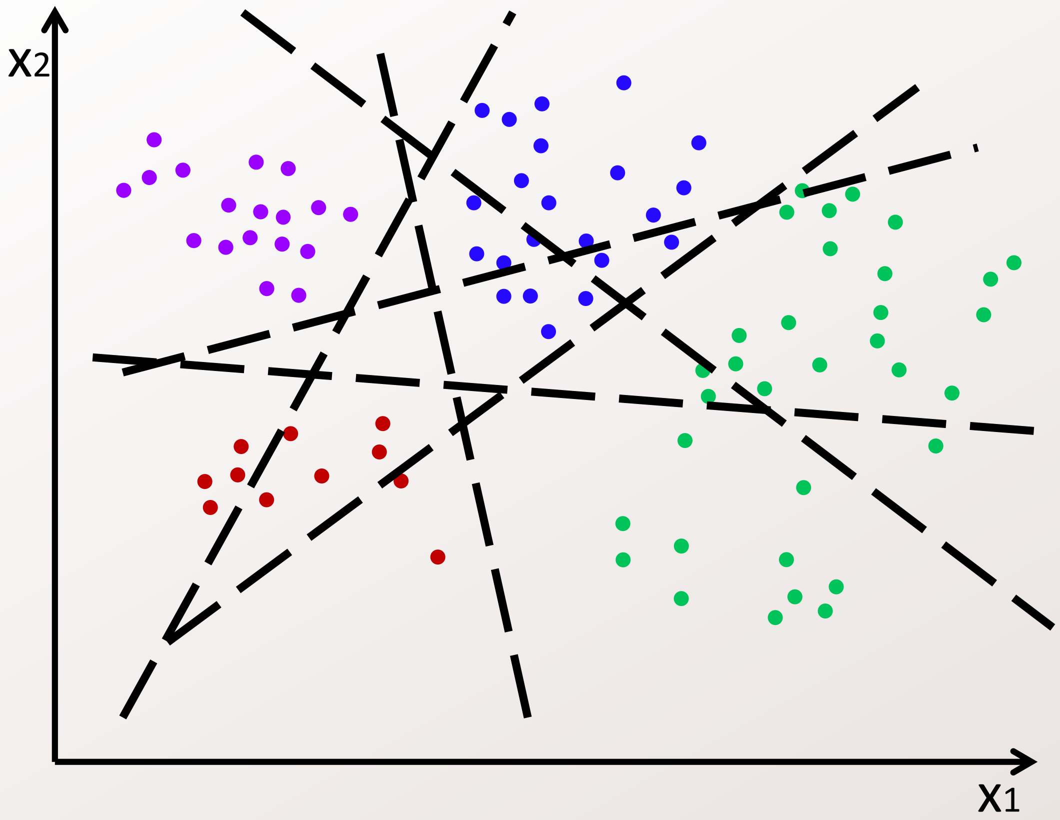
A decision tree



Toy tree learning example

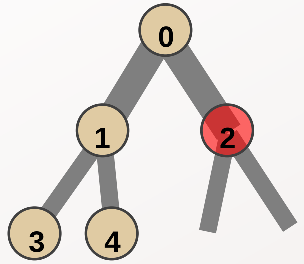


- Try several hyperplanes, chosen at random
- Keep hyperplane that best separates data
  - information gain
- Recurse

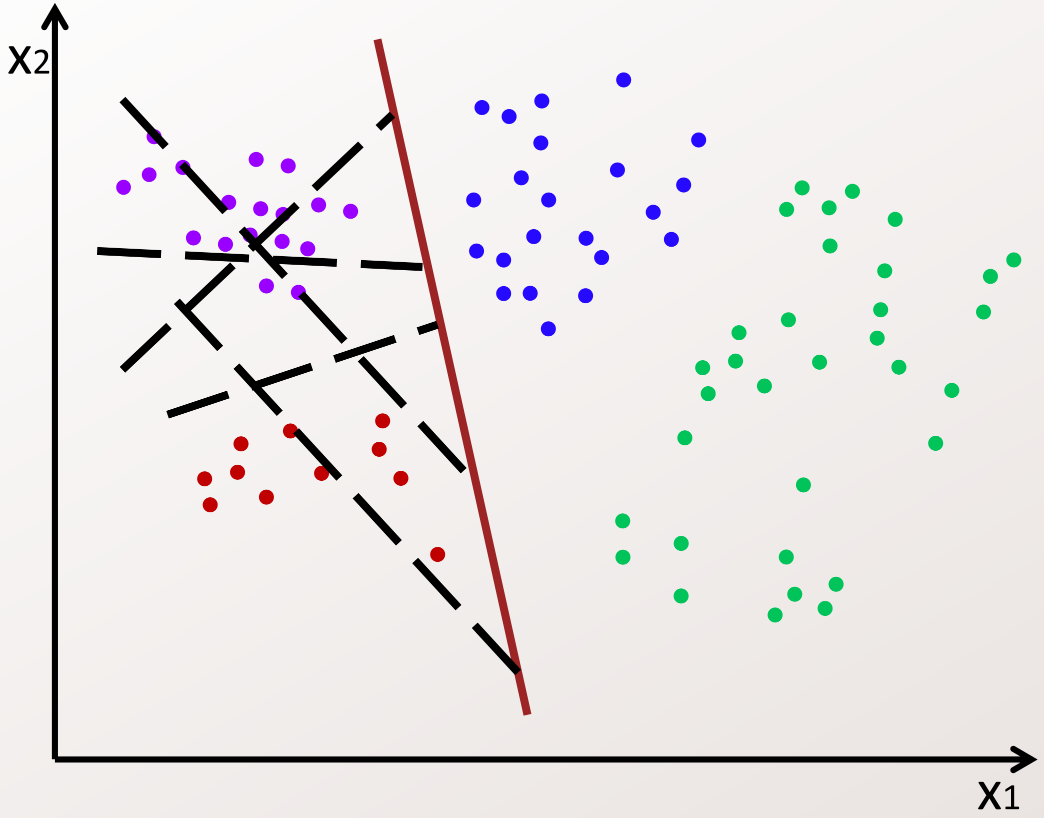


Learning a conditional structure of discriminative features.

Toy tree learning example

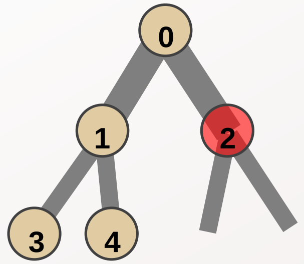


- Try several hyperplanes, chosen at random
- Keep hyperplane that best separates data
  - information gain
- Recurse

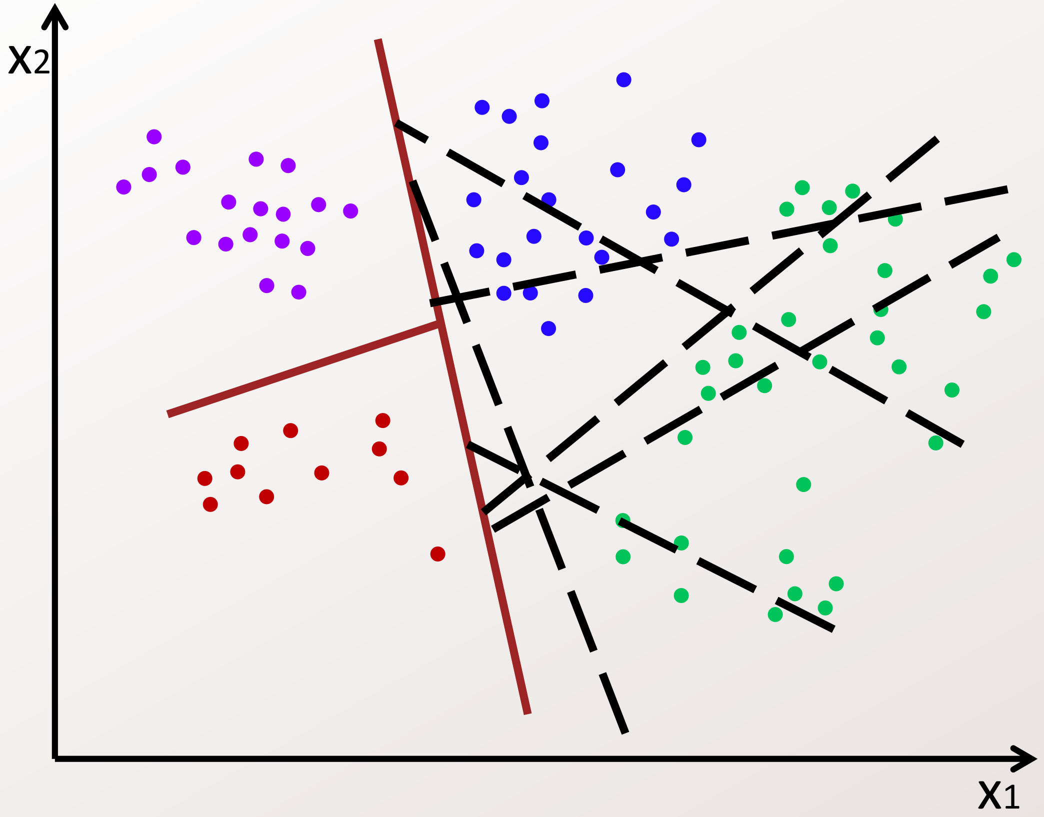


Learning a conditional structure of discriminative features.

Toy tree learning example

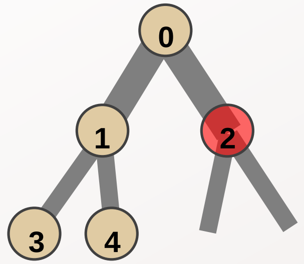


- Try several hyperplanes, chosen at random
- Keep hyperplane that best separates data
  - information gain
- Recurse

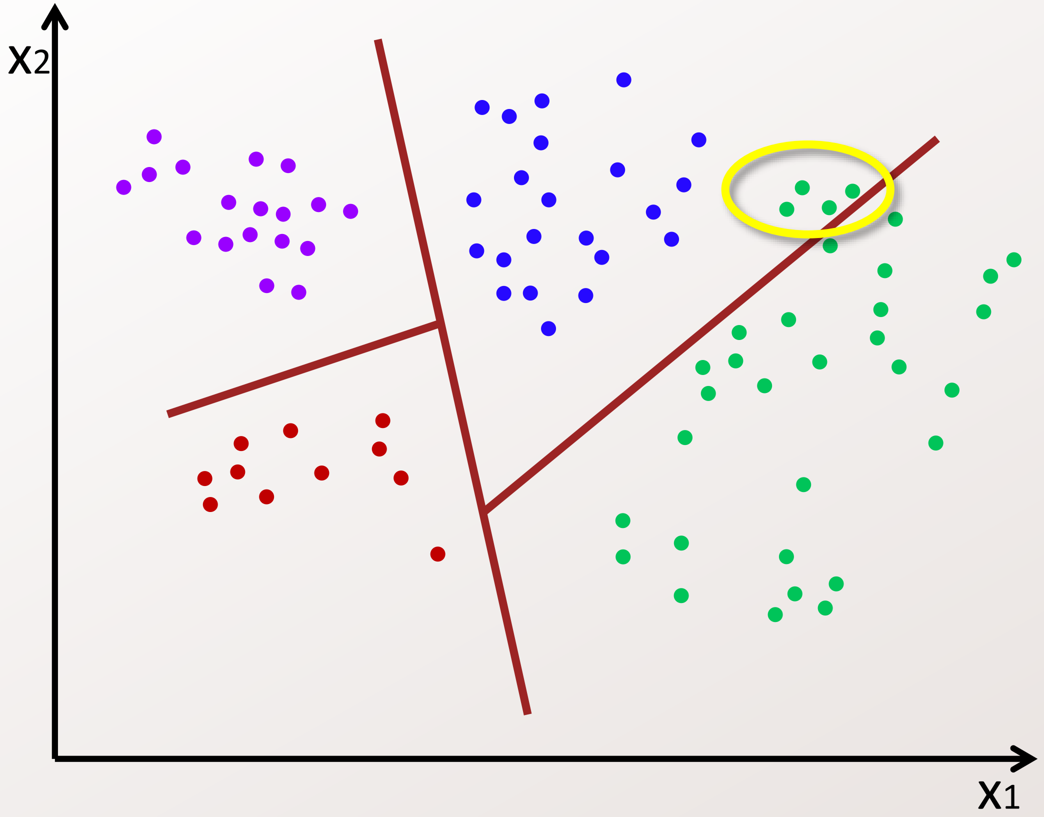


Learning a conditional structure of discriminative features.

Toy tree learning example

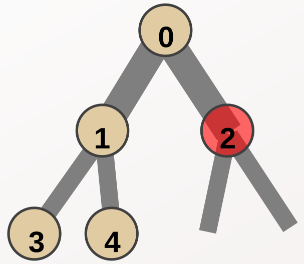


- Try several hyperplanes, chosen at random
- Keep hyperplane that best separates data
  - information gain
- Recurse



Learning a conditional structure of discriminative features.

Training objective function

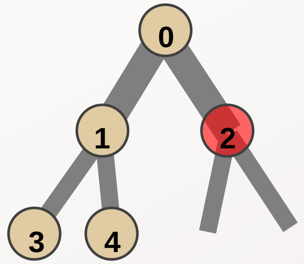


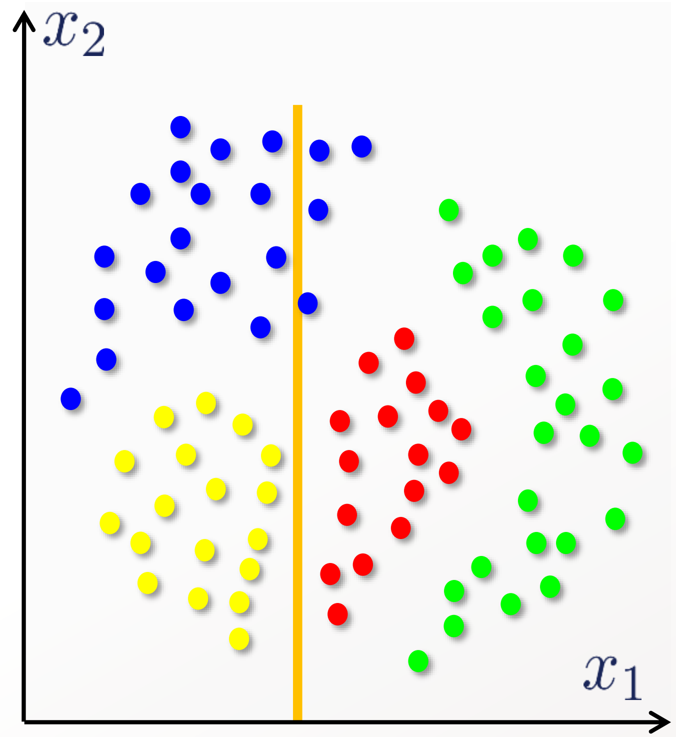
- Used to decide which candidate **split function** is best

- Typically an “information gain” – a very general and flexible formulation

$$I = \underbrace{H(S_j)}_{\text{entropy of examples at parent node}} - \sum_{i=\mathrm{L},\mathrm{R}} \underbrace{\frac{|S_j^i|}{|S_j|}}_{\text{weighting left/right children}} \underbrace{H(S_j^i)}_{\text{entropy of examples at child nodes}}$$

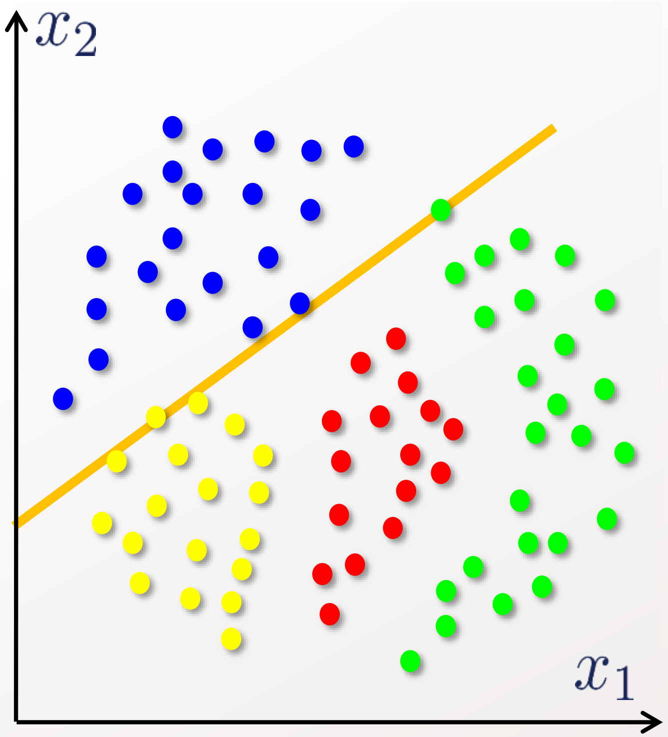
Examples of split functions



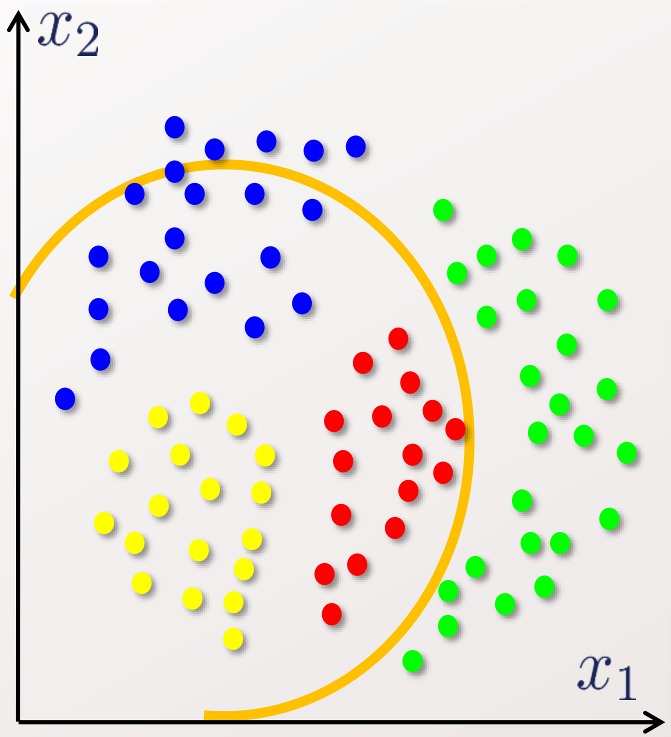


“Axis aligned”

Efficient (one feature at a time)

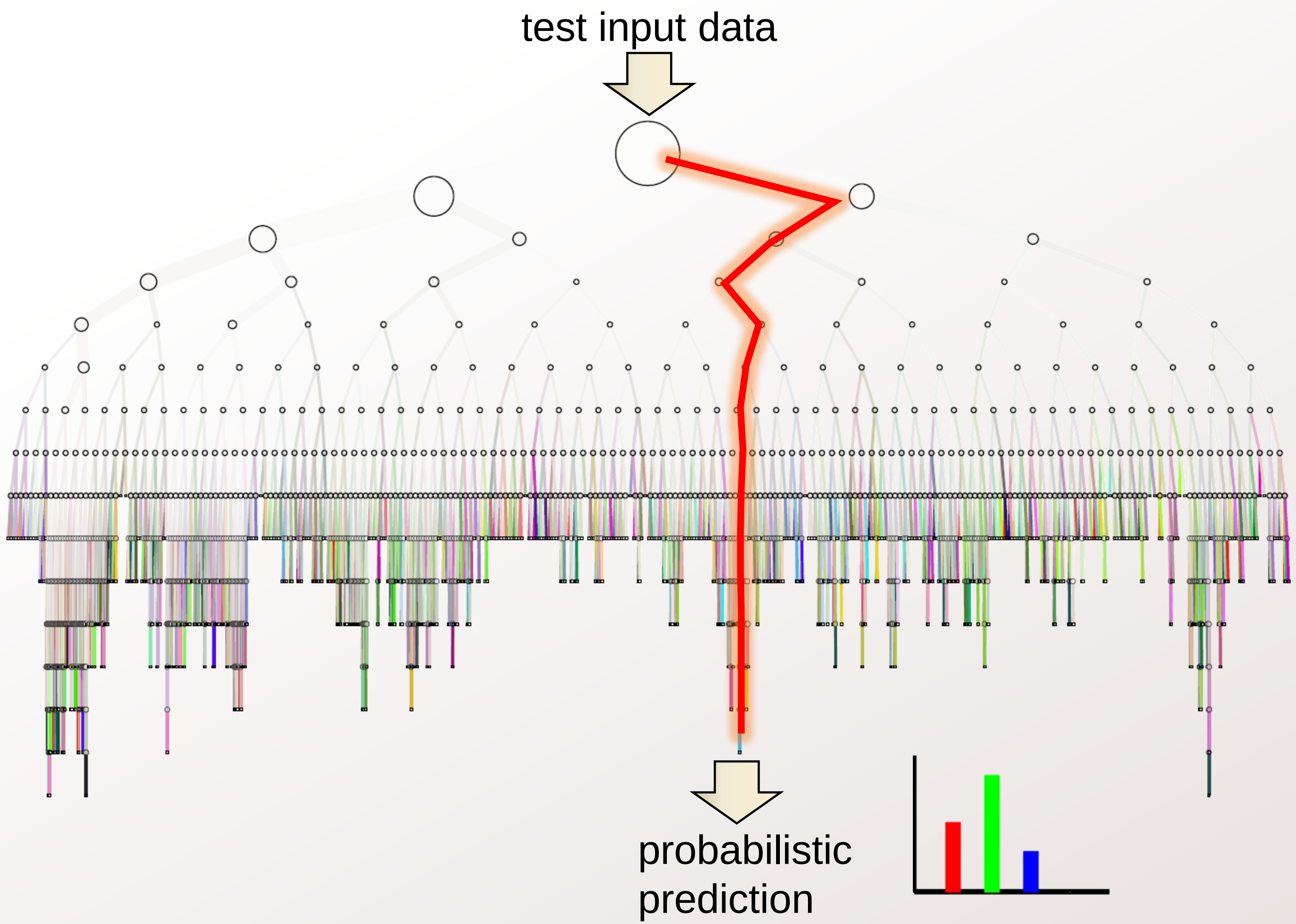


“Oriented hyper-plane”

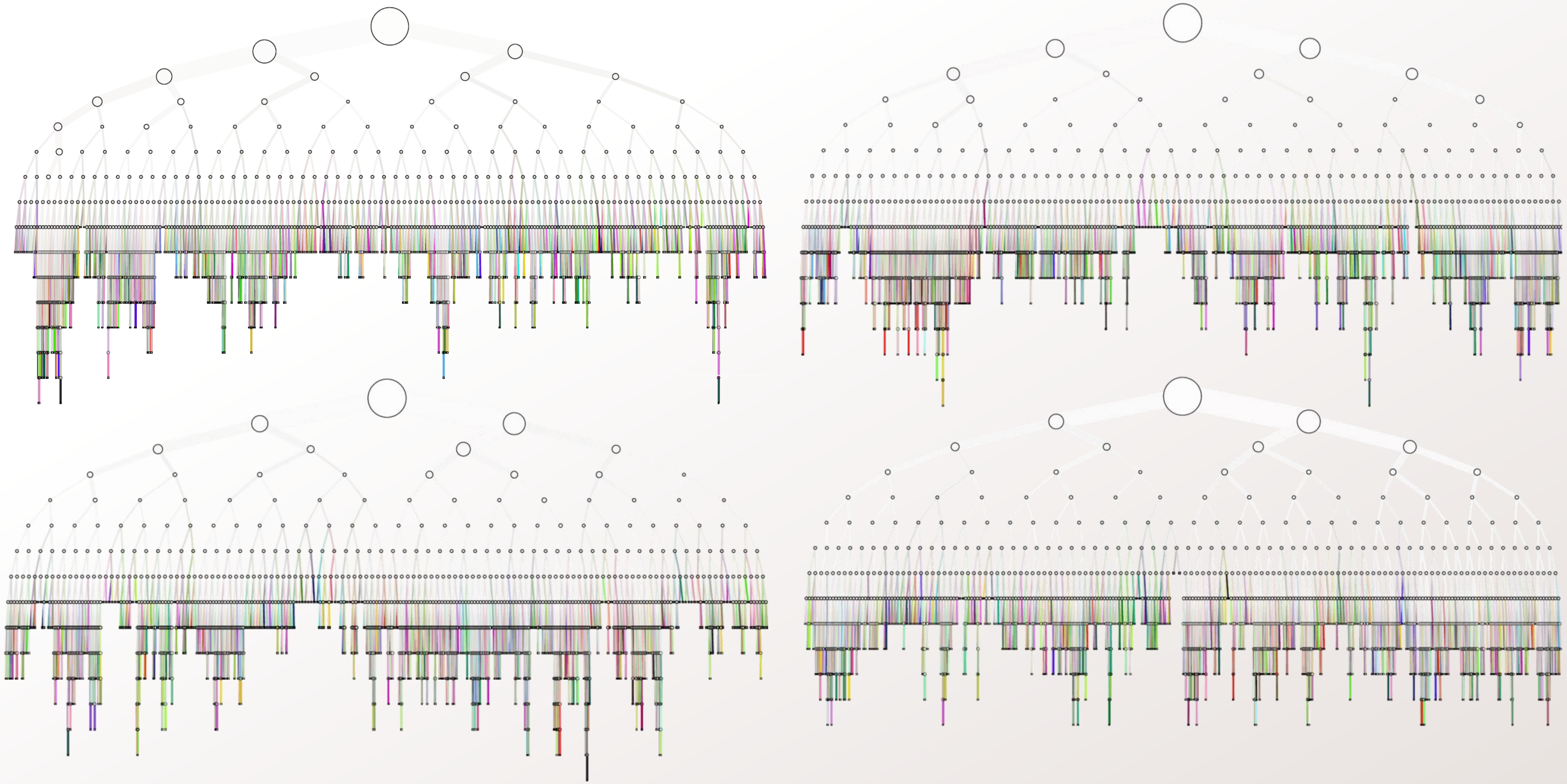


“Conic section”

Decision trees: test time prediction

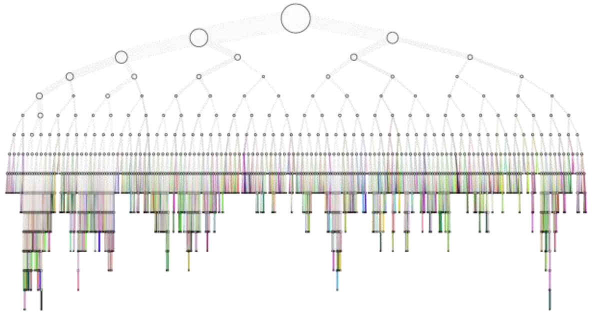


Decision forests

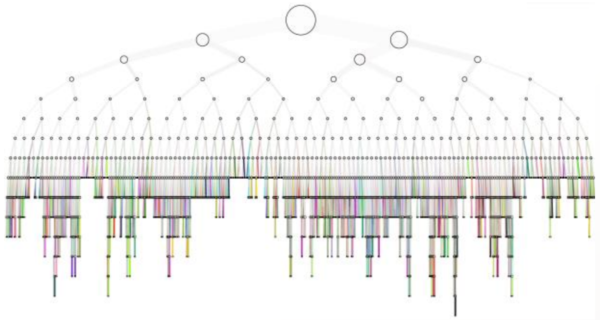


Better generalization than individual trees

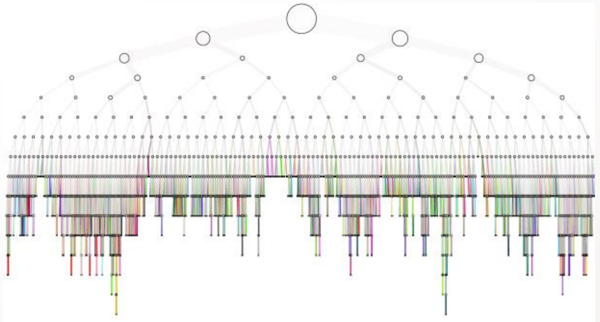
Aggregating tree predictions



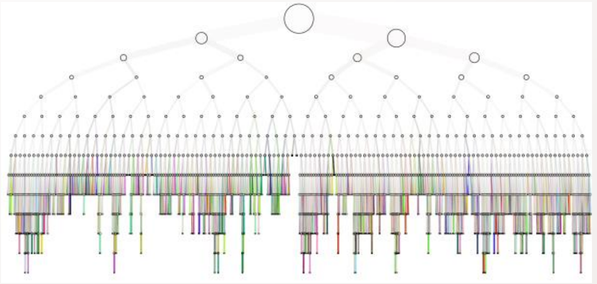
$p_{t=1}(y|\mathbf{v})$



$p_{t=2}(y|\mathbf{v})$

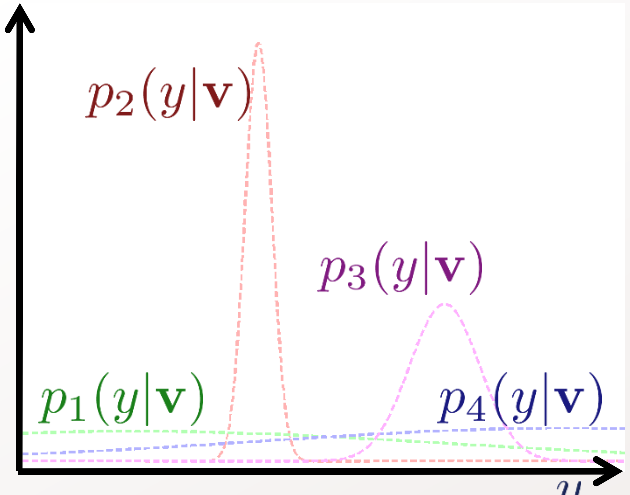


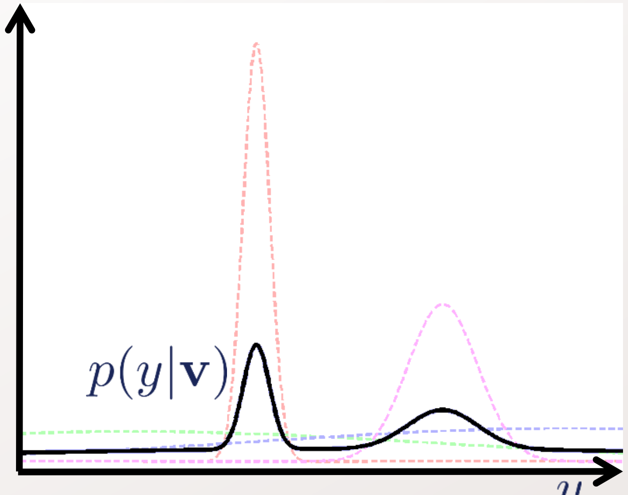
$p_{t=3}(y|\mathbf{v})$

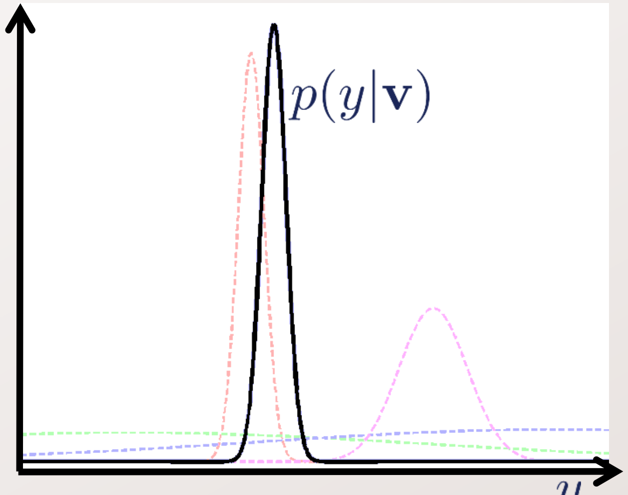


$p_{t=4}(y|\mathbf{v})$

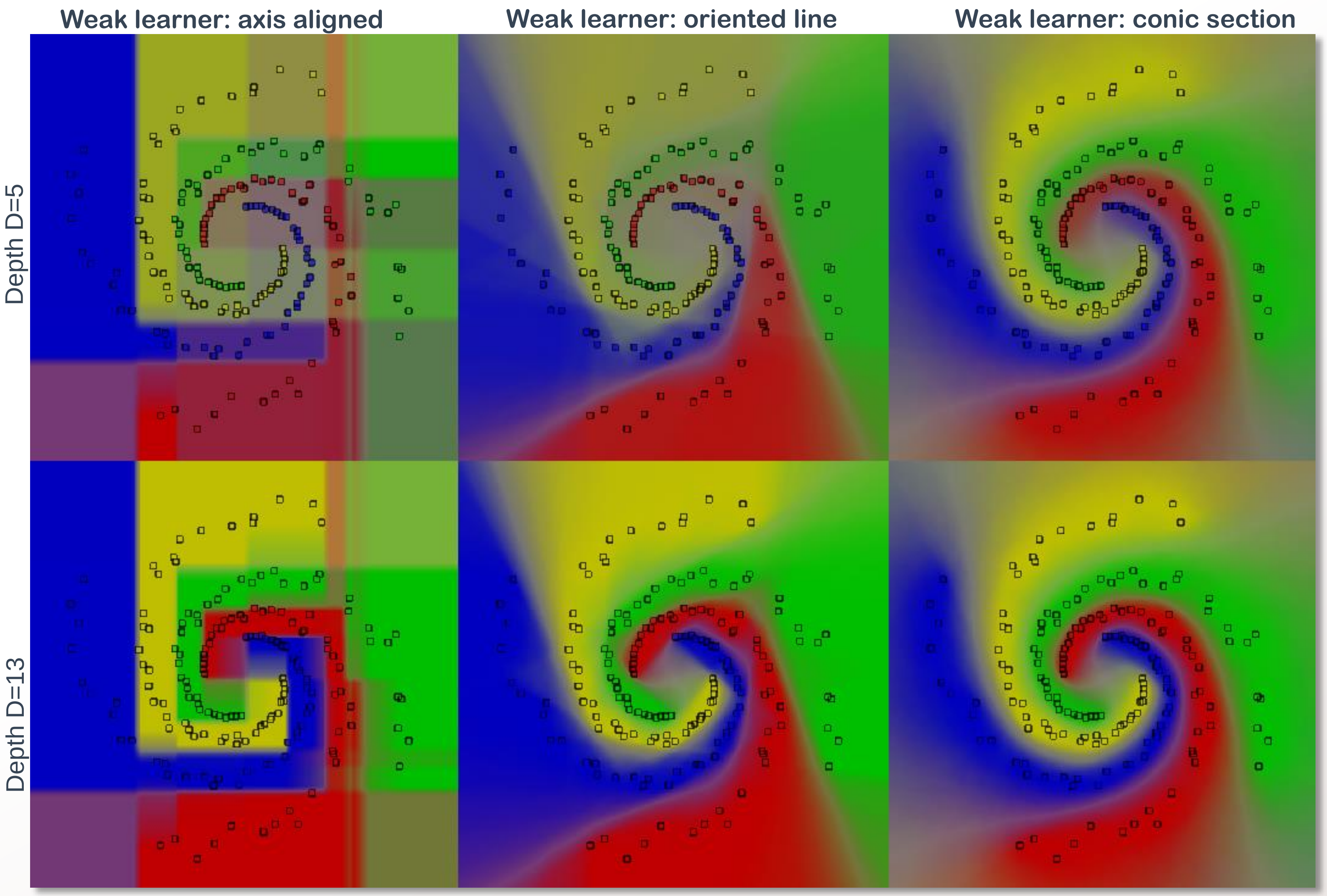
$$p(y|\mathbf{v}) = \frac{1}{T}\sum_{t}^{T} p_t(y|\mathbf{v}) \qquad p(y|\mathbf{v}) = \frac{1}{Z}\prod_{t}^{T} p_t(y|\mathbf{v})$$

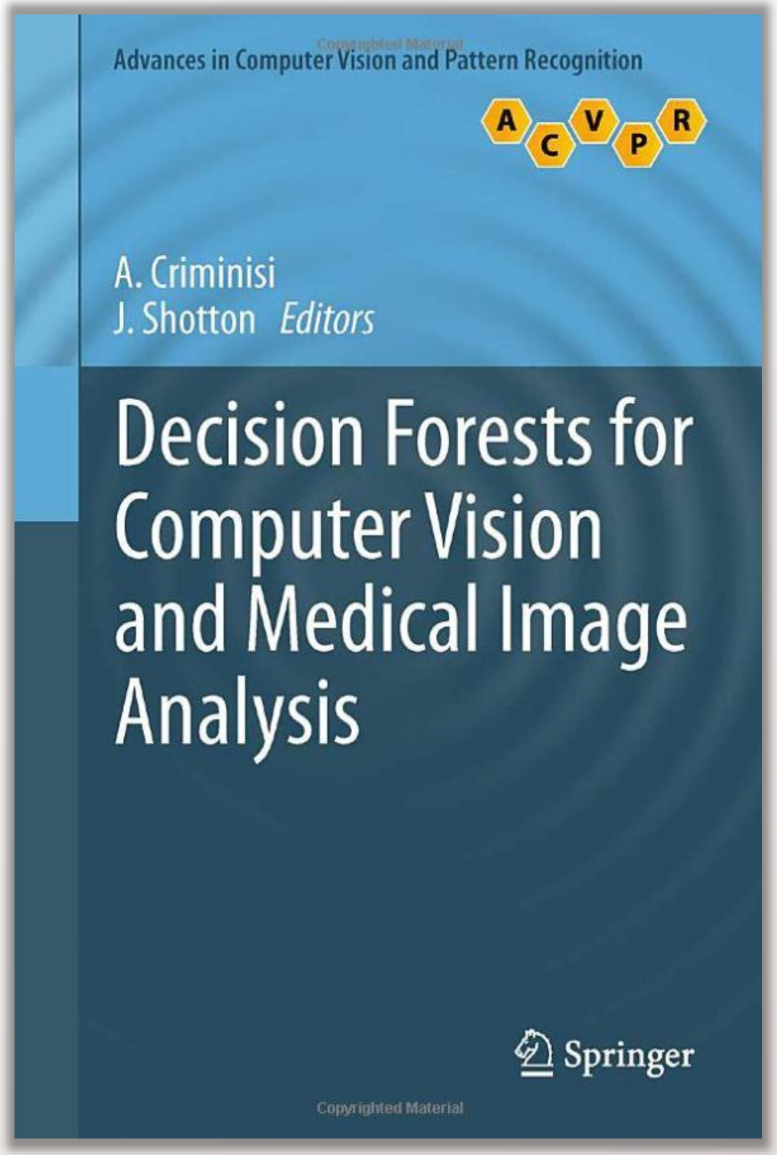






Effect of tree depth and randomness

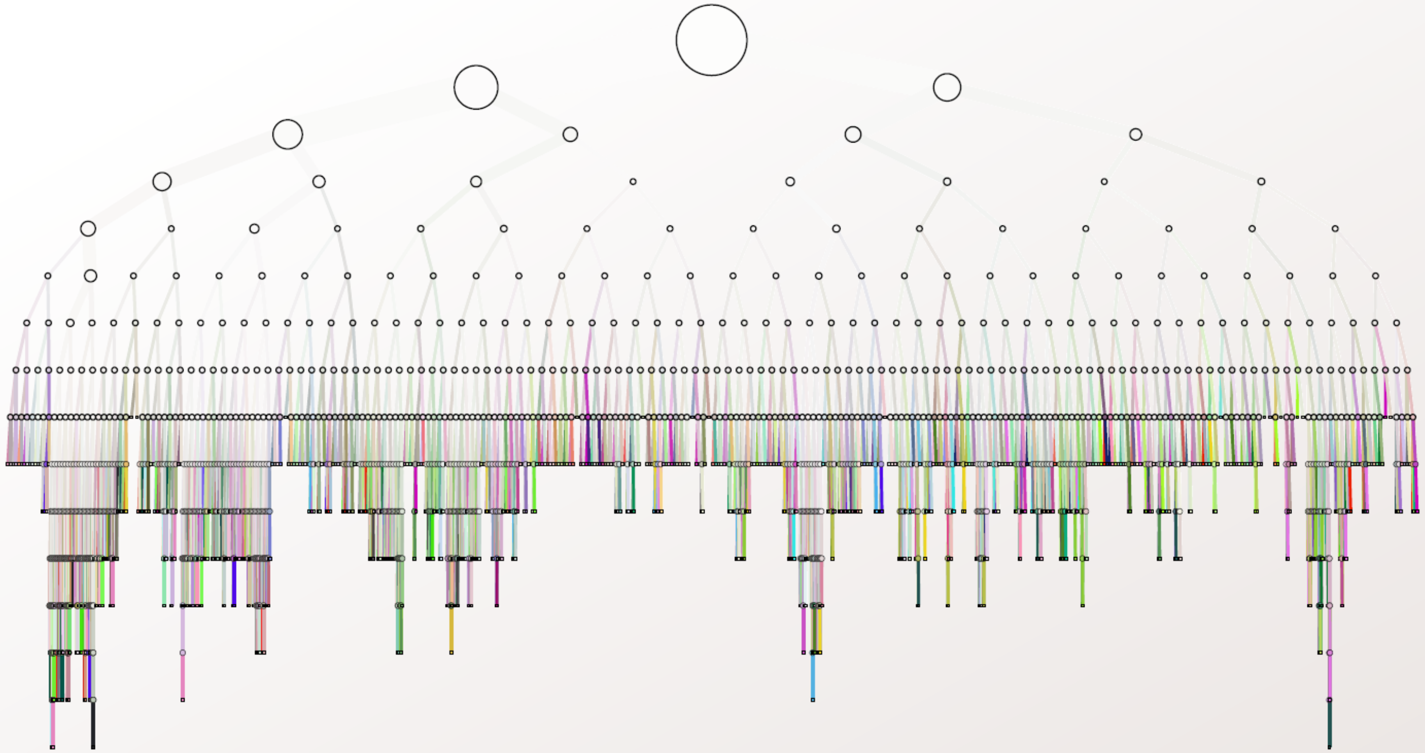




Free code available!

A. Criminisi and J. Shotton, **Decision Forests for Computer Vision and Medical Image Analysis**, Springer, February 2013

Are forests sufficient?

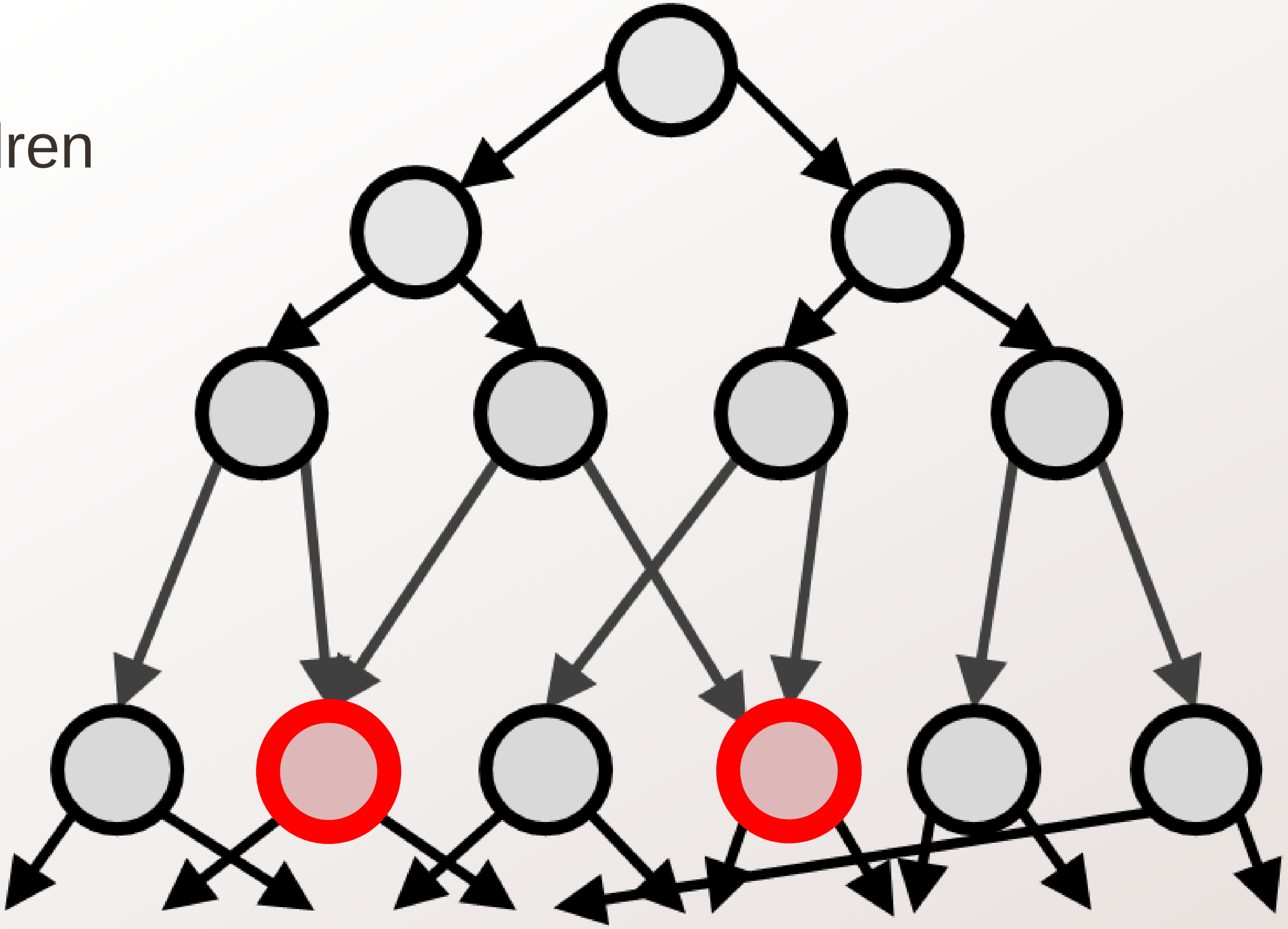


- Memory issues:
  - Number of nodes in trees
    grows exponentially with depth

- Amount of training data
  - Training data is quickly diluted with depth
  - Yet, training deeper trees (on enough data) yields highest test accuracy
    (several real applications, e.g. Kinect, have “infinite” data available)

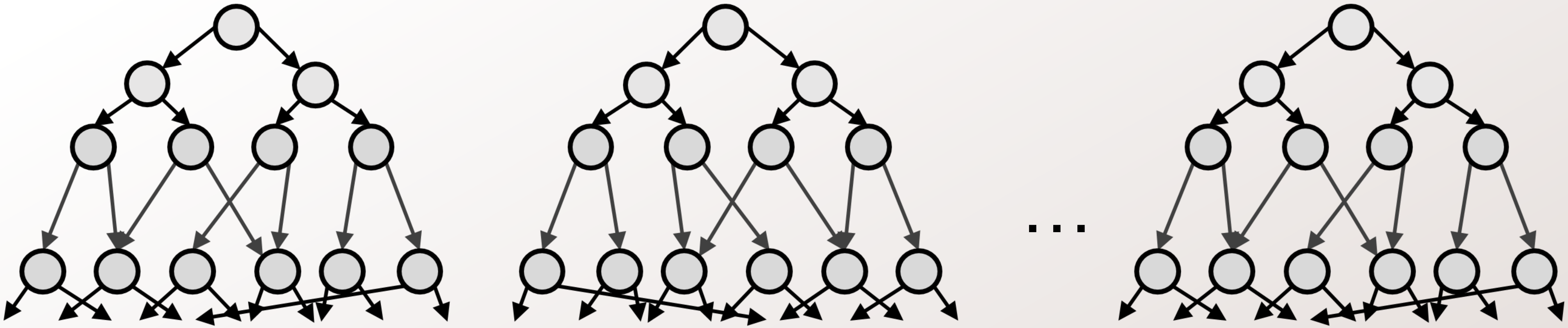
From trees to DAGs: node merging

- Each internal node has 2 children (like in binary trees)
- Each non-root node can have more than 1 parent



Decision jungles

- A “jungle” is an ensemble of *rooted* decision DAGs
- We train each DAG layer by layer, jointly optimizing both
  - the structure of the DAG
  - the split node features



Properties of jungles

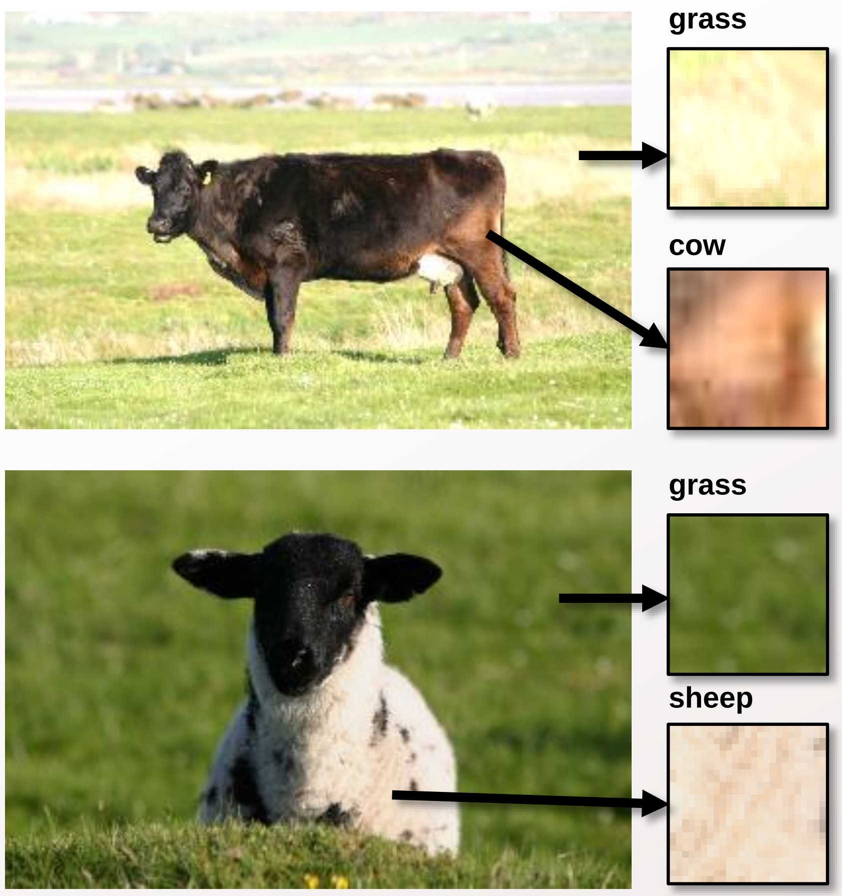
- **Limited memory consumption**
  - e.g. by specifying a width at each layer in the DAG

- Potentially improved **generalization**
  - fewer parameters
  - **less “dilution”** of training data

How do DAGs help in practice?

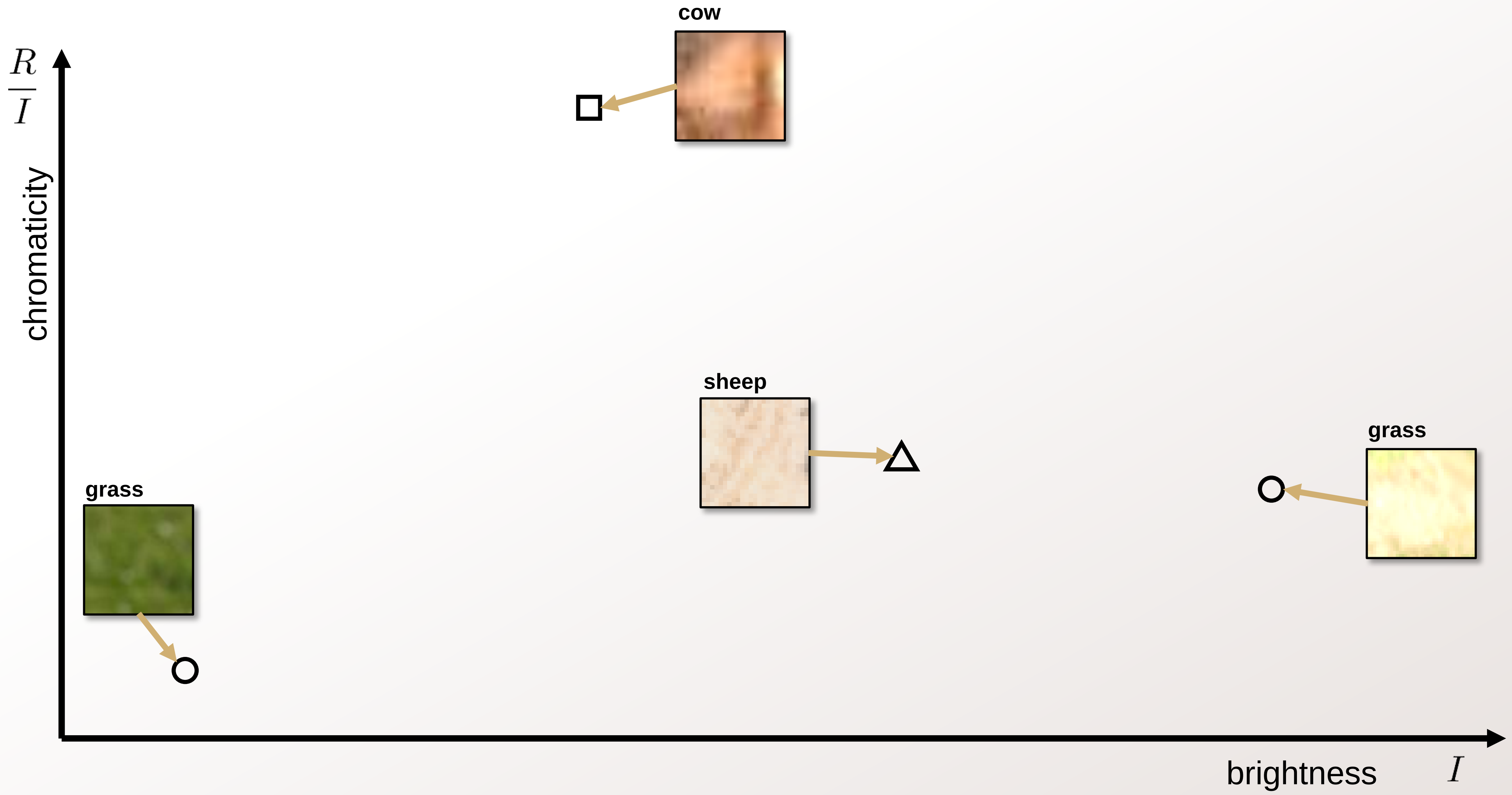
A toy example on classifying images of cows, sheep and grass

Training data

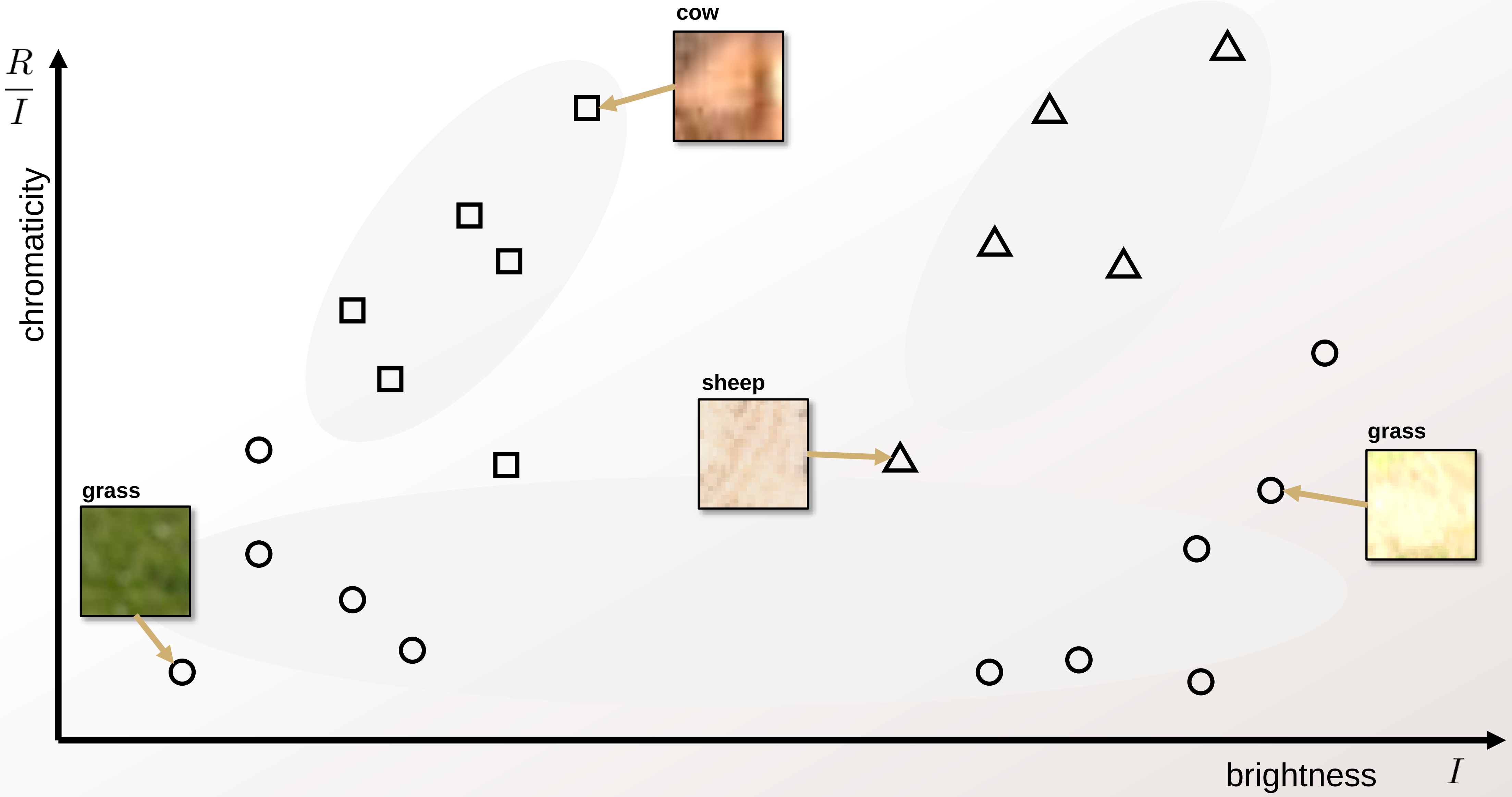


...

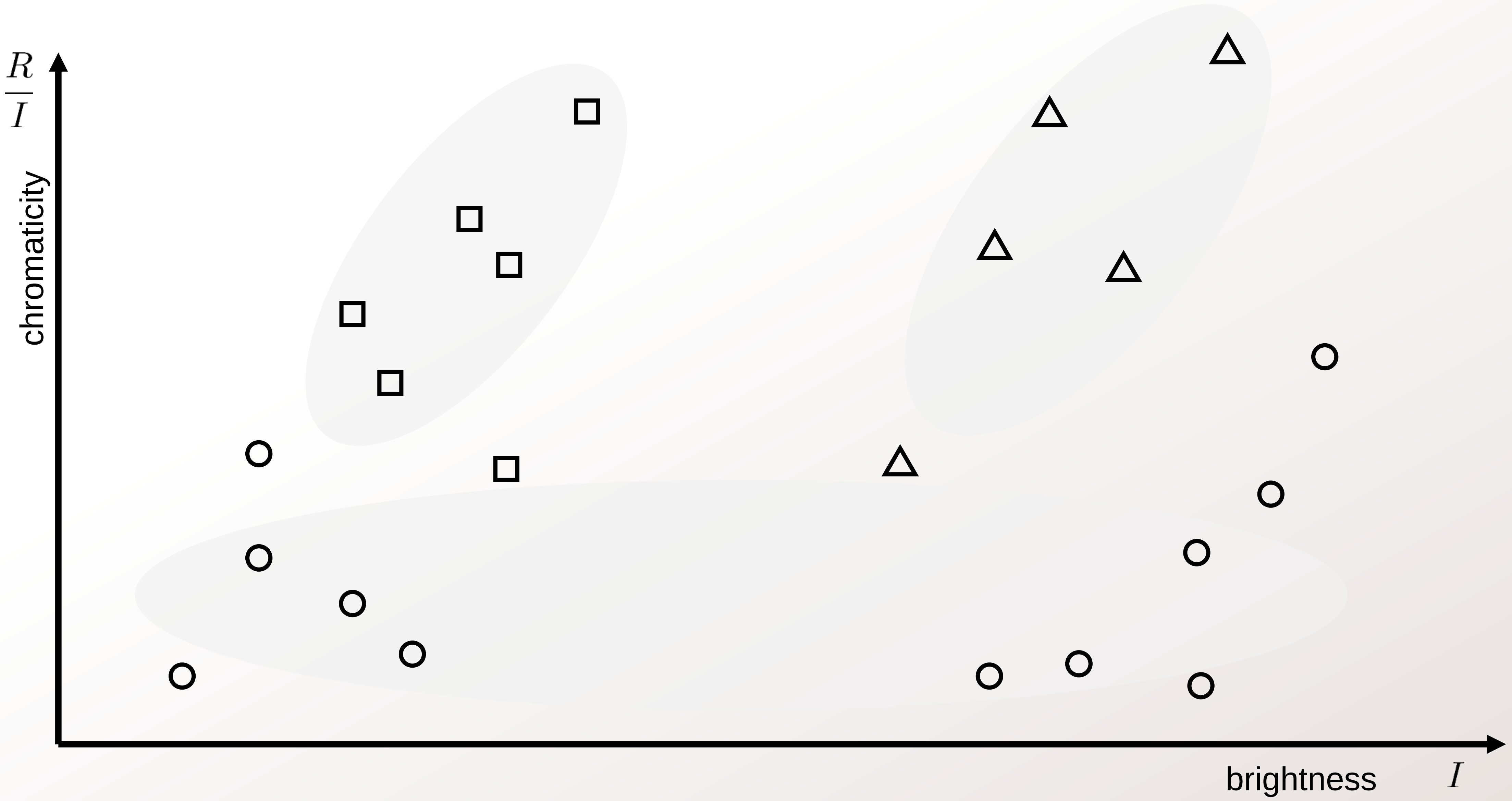
A toy example on classifying images of cows, sheep and grass



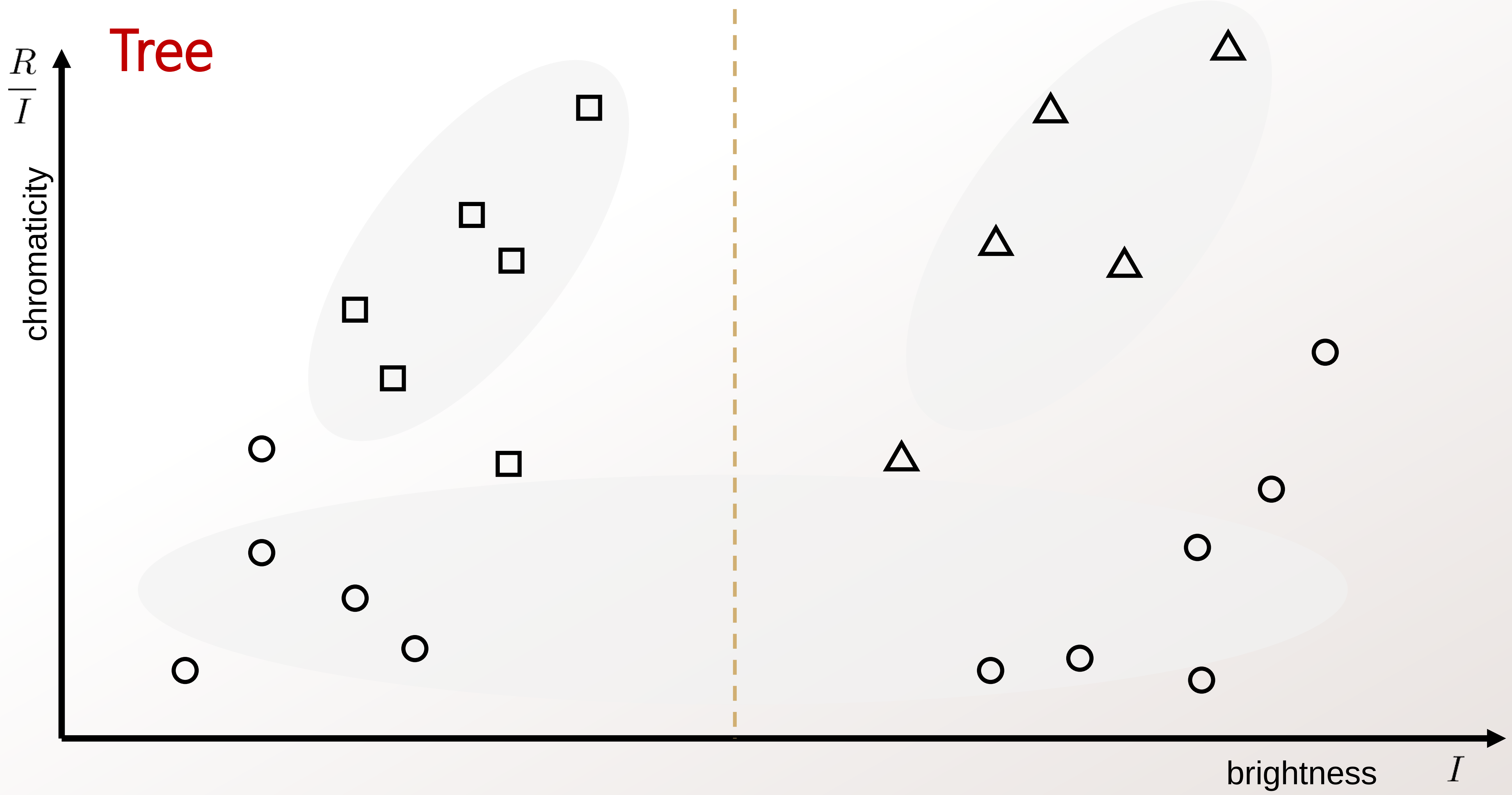
A toy example on classifying images of cows, sheep and grass



A toy example on classifying images of cows, sheep and grass

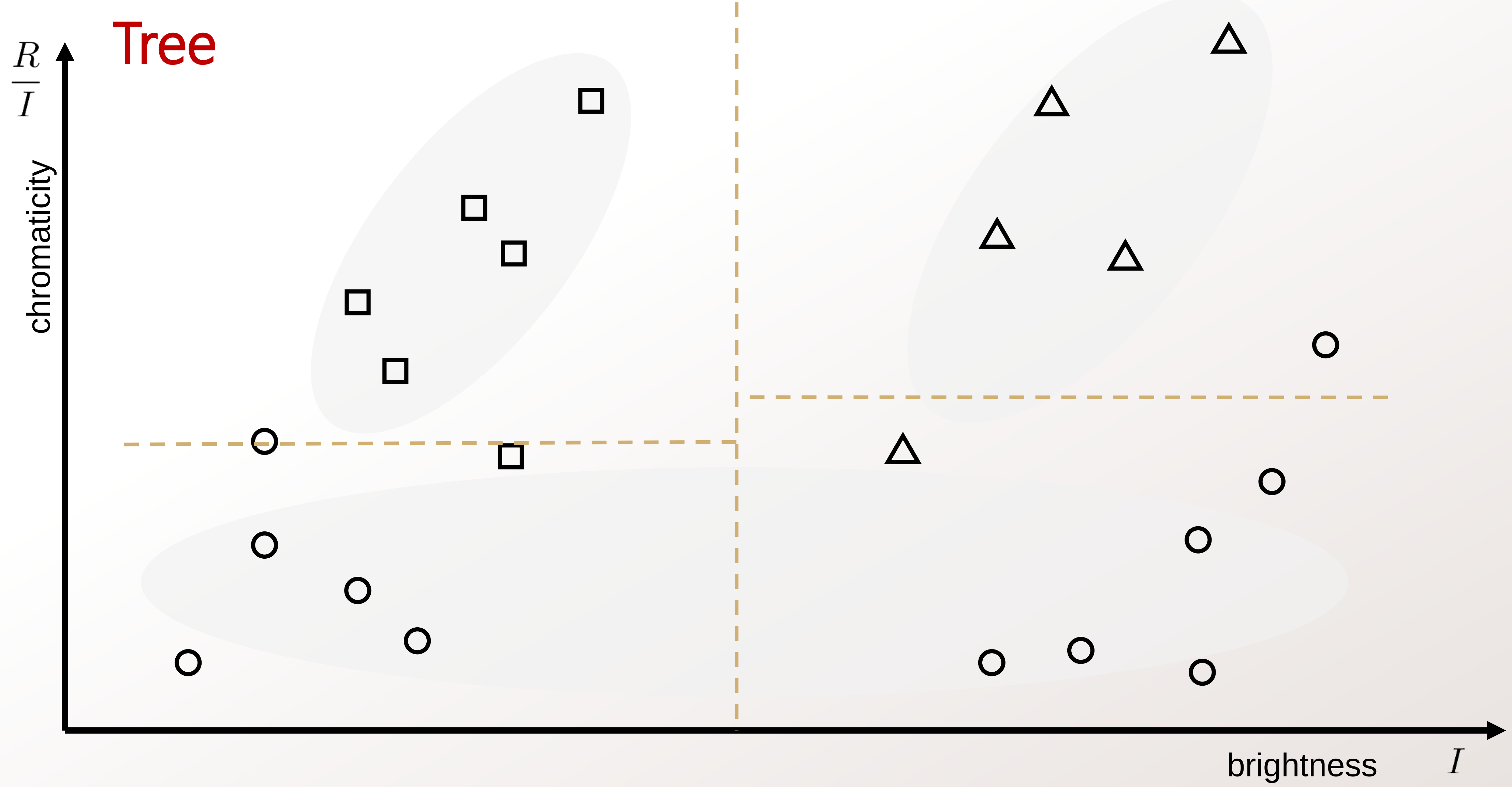


A toy example on classifying images of cows, sheep and grass

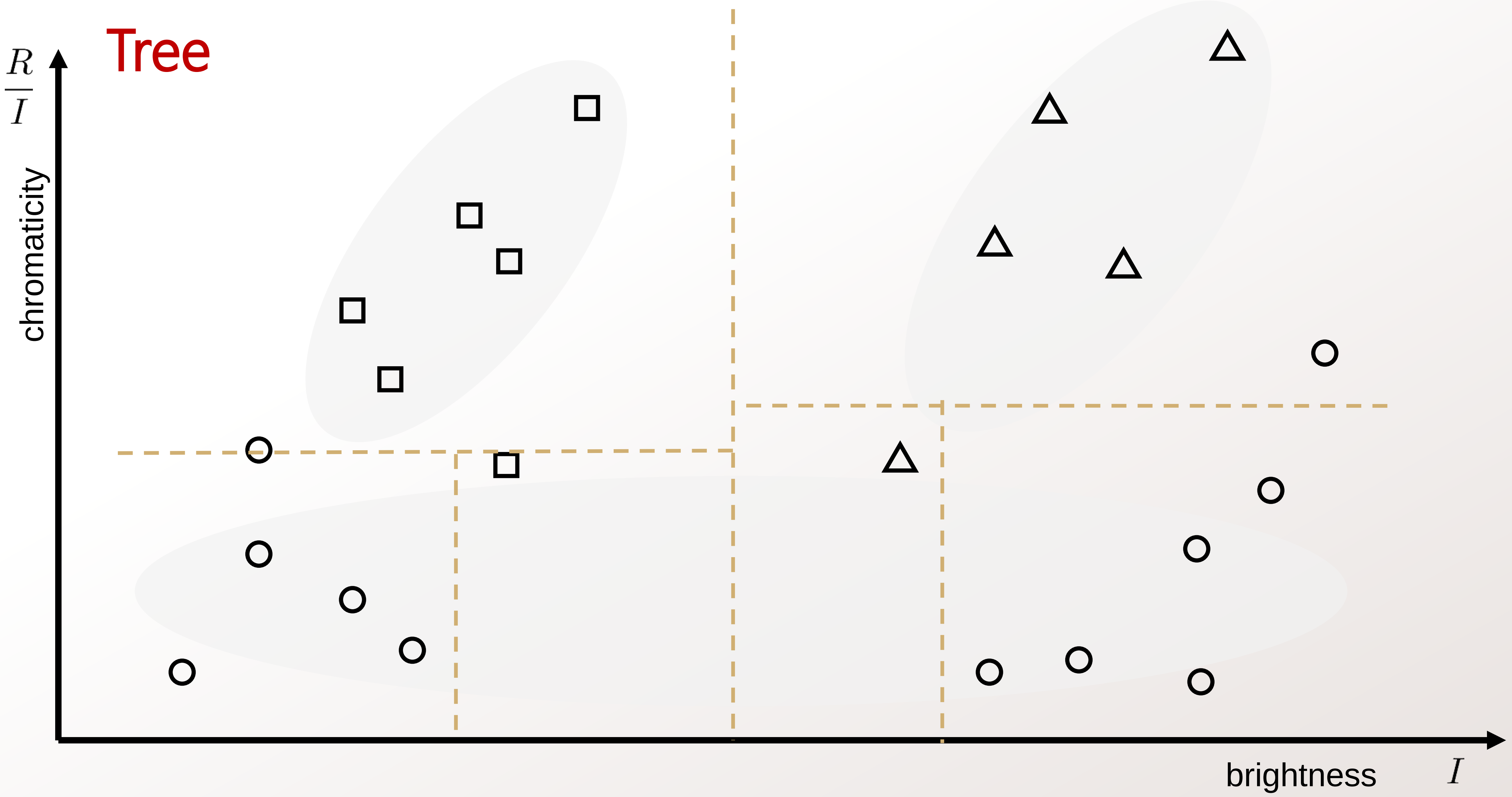


Axis-aligned splits only

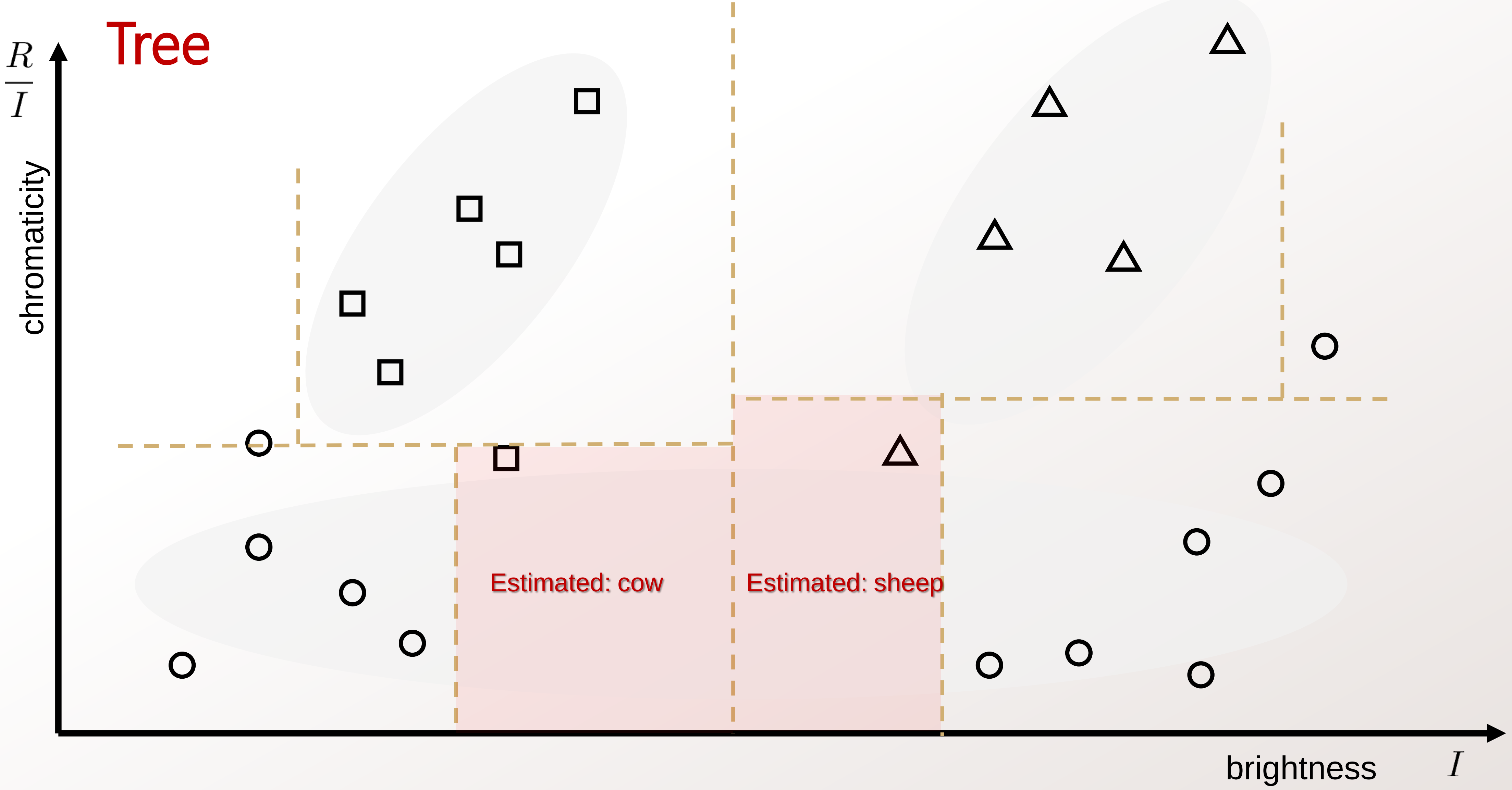
A toy example on classifying images of cows, sheep and grass



A toy example on classifying images of cows, sheep and grass

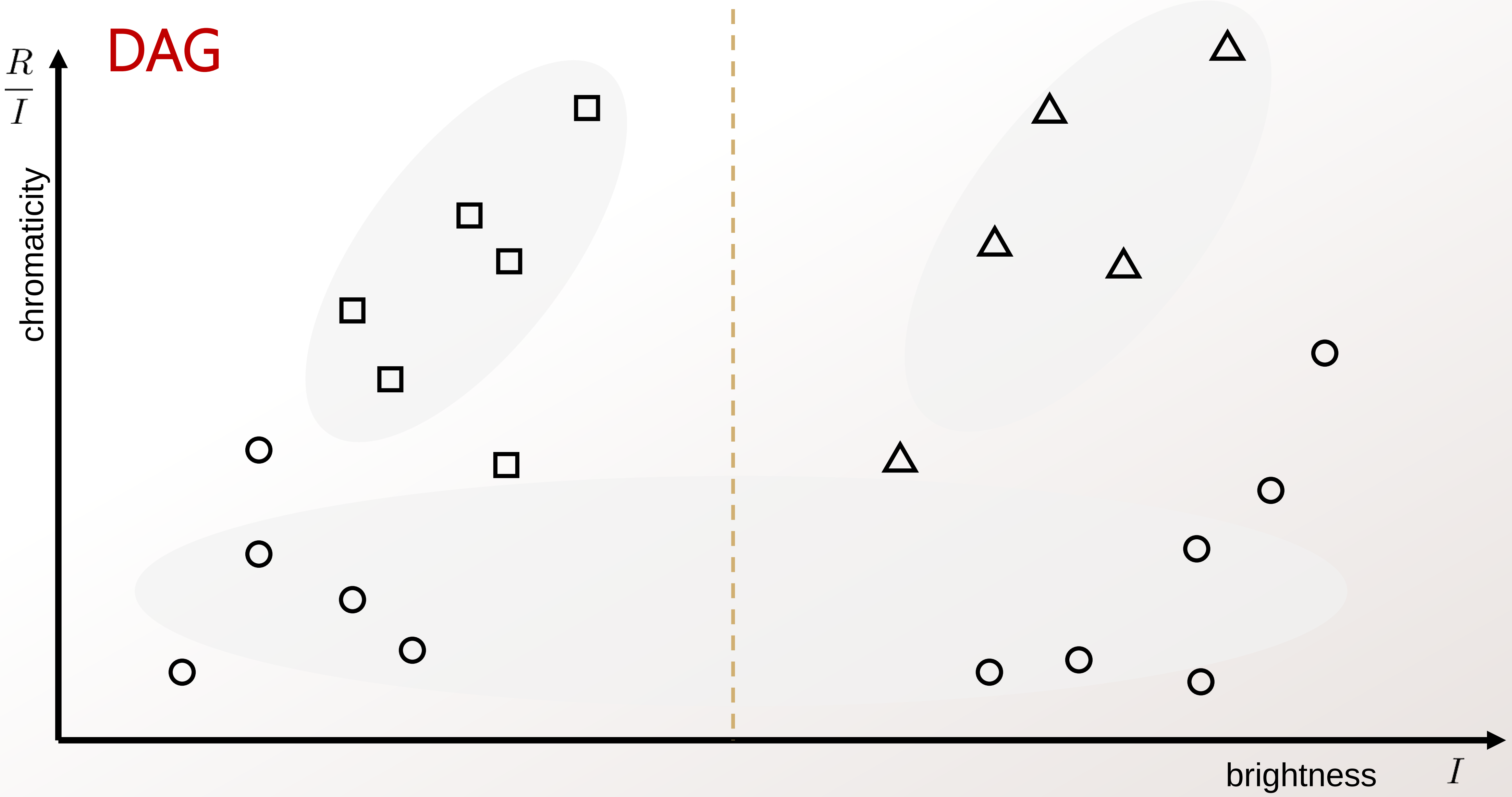


A toy example on classifying images of cows, sheep and grass

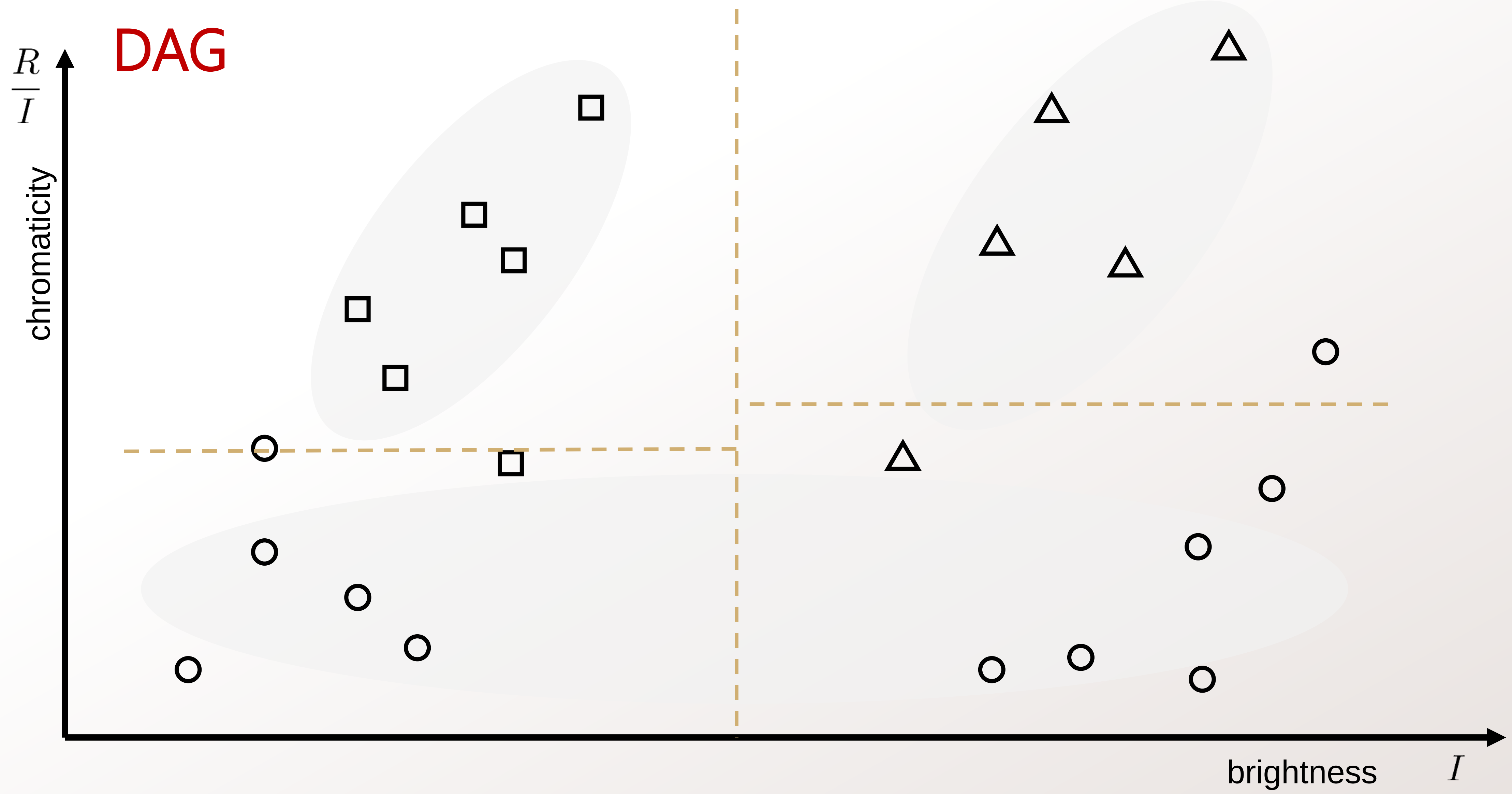


Too many model parameters: overfitting

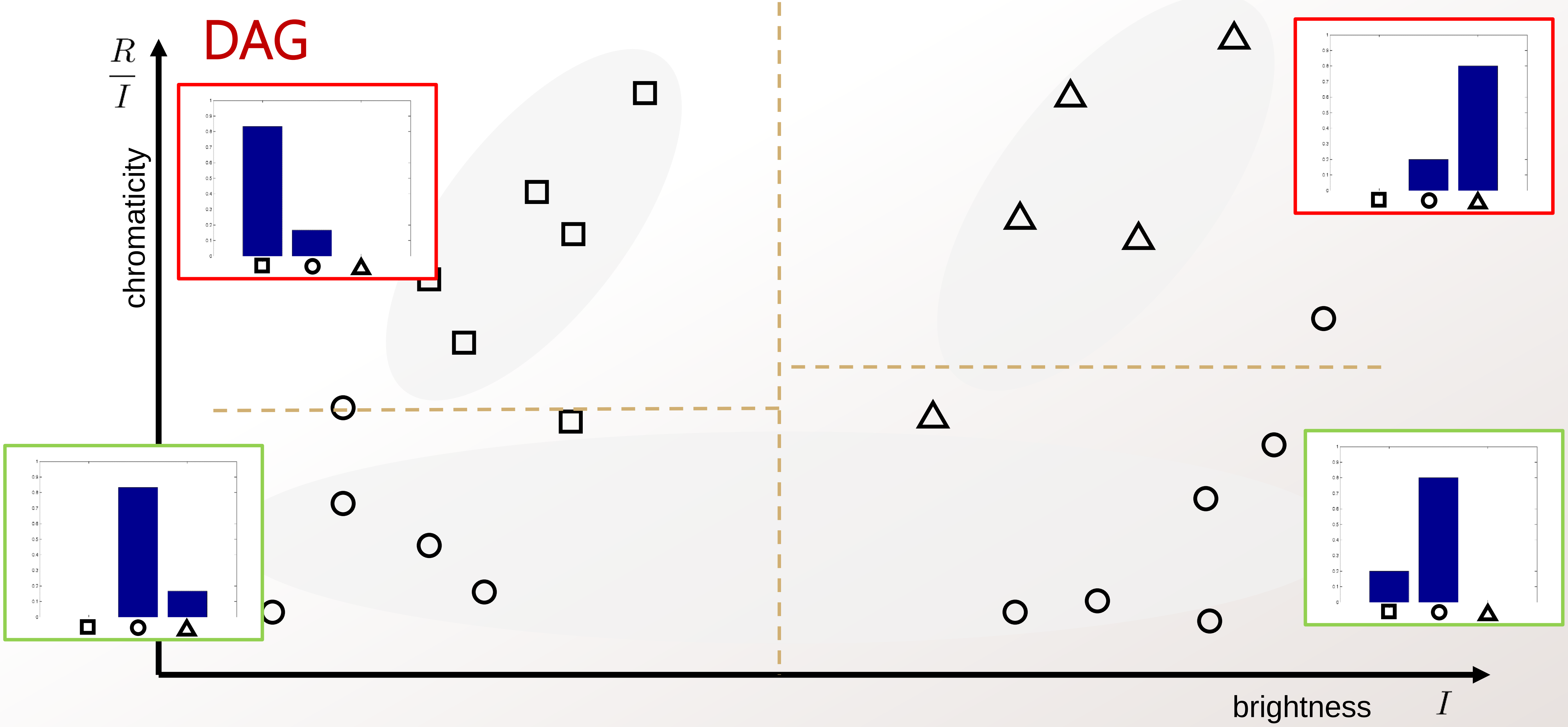
A toy example on classifying images of cows, sheep and grass



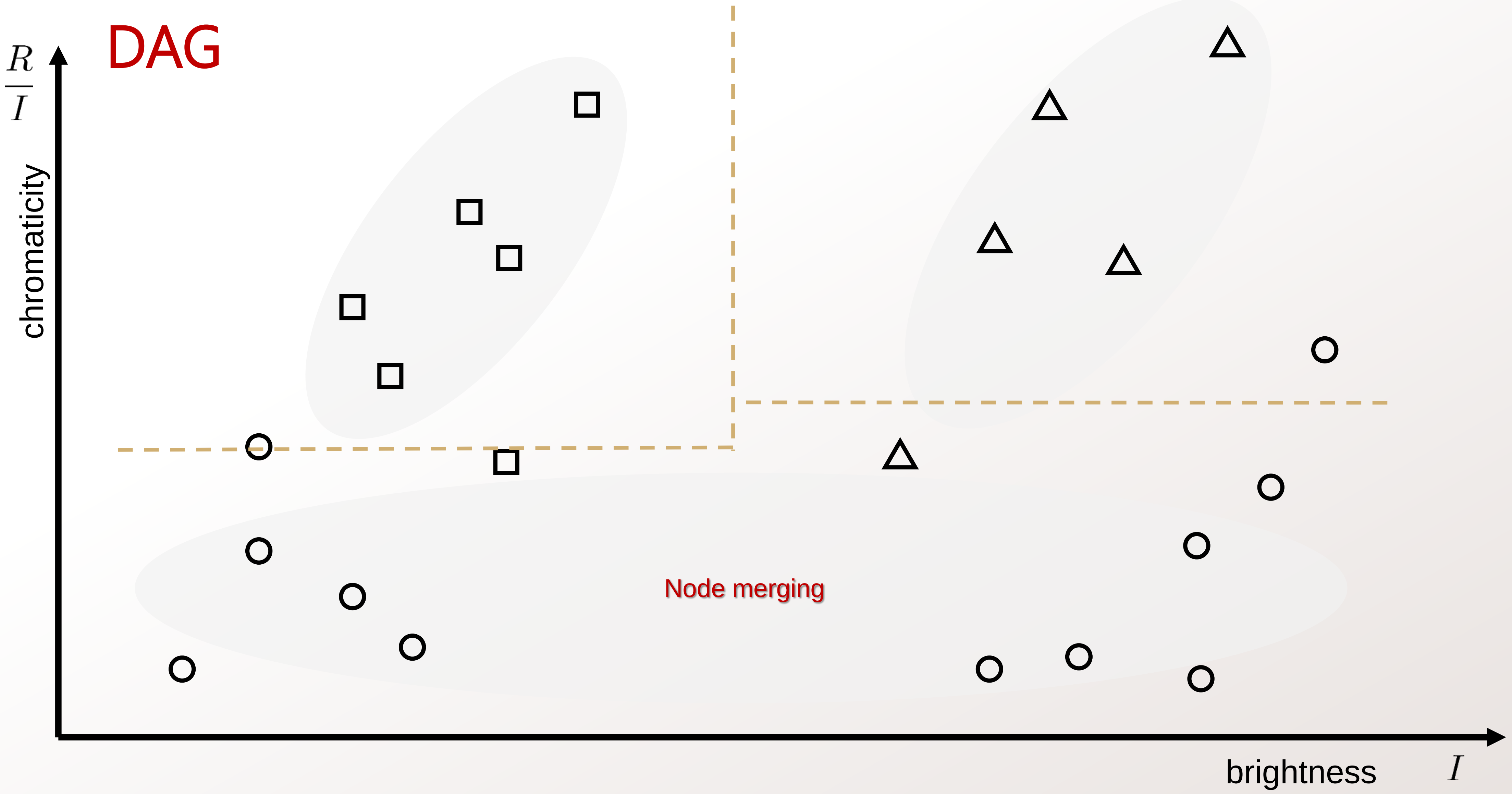
A toy example on classifying images of cows, sheep and grass



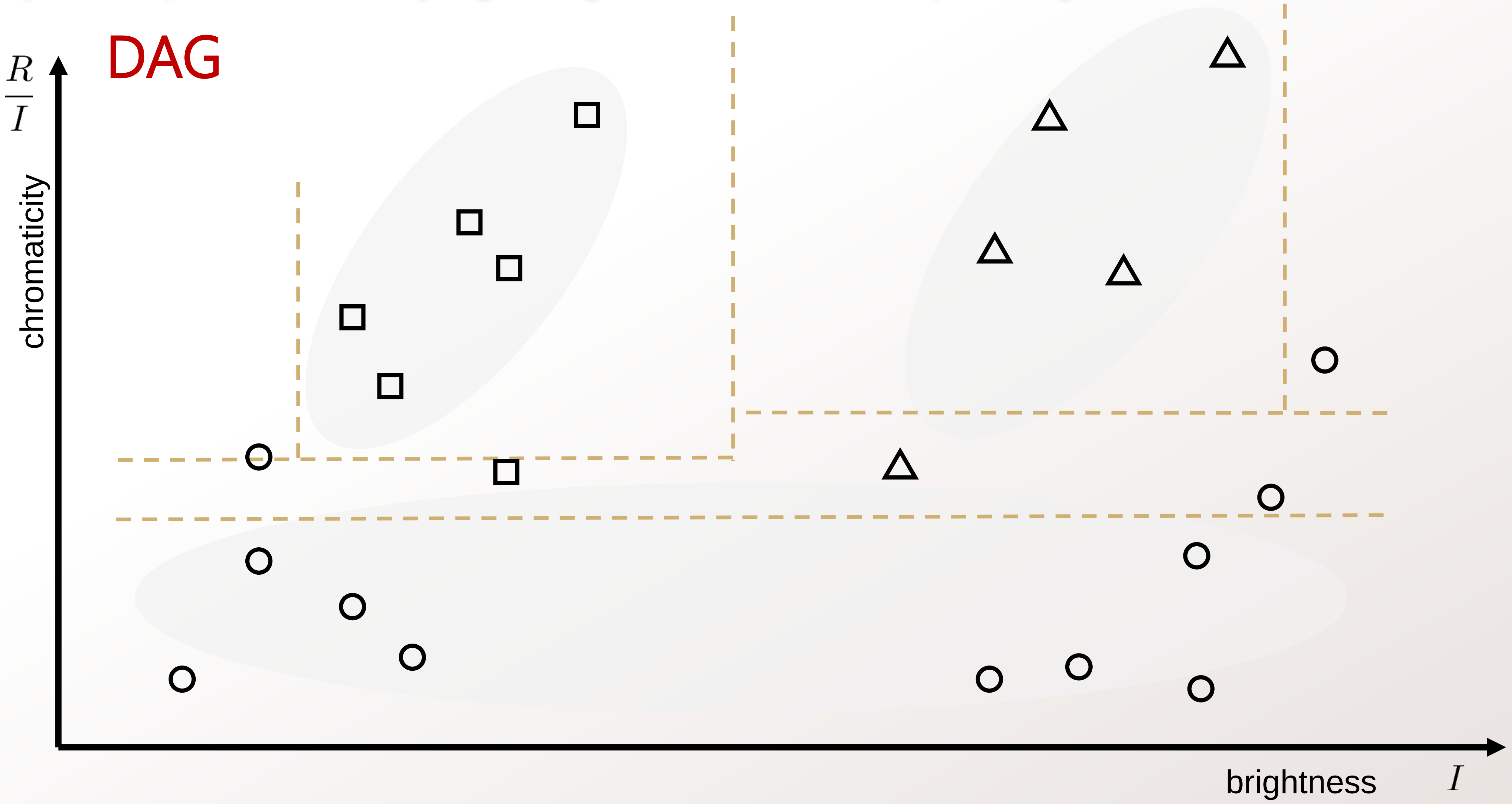
A toy example on classifying images of cows, sheep and grass



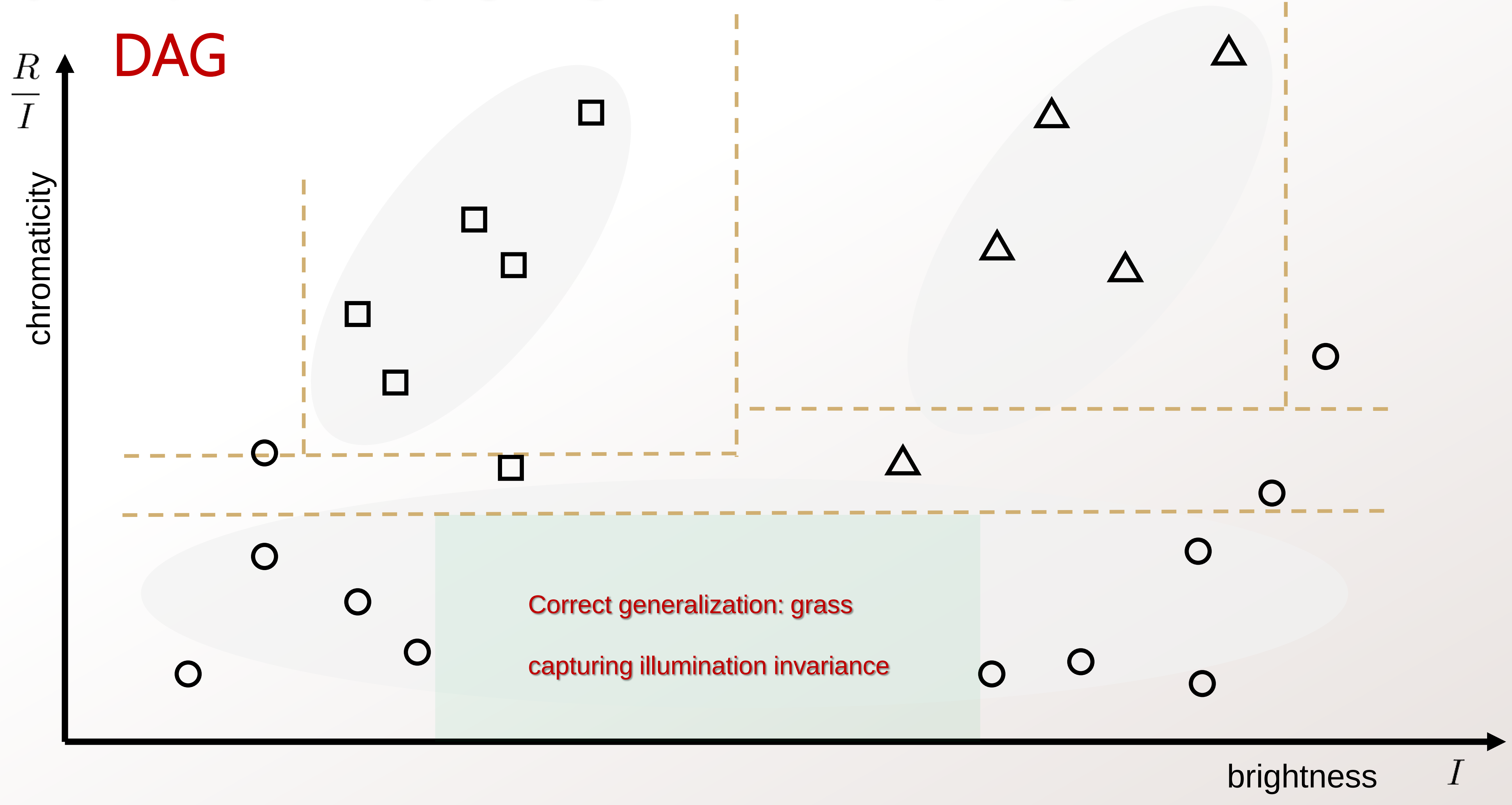
A toy example on classifying images of cows, sheep and grass



A toy example on classifying images of cows, sheep and grass

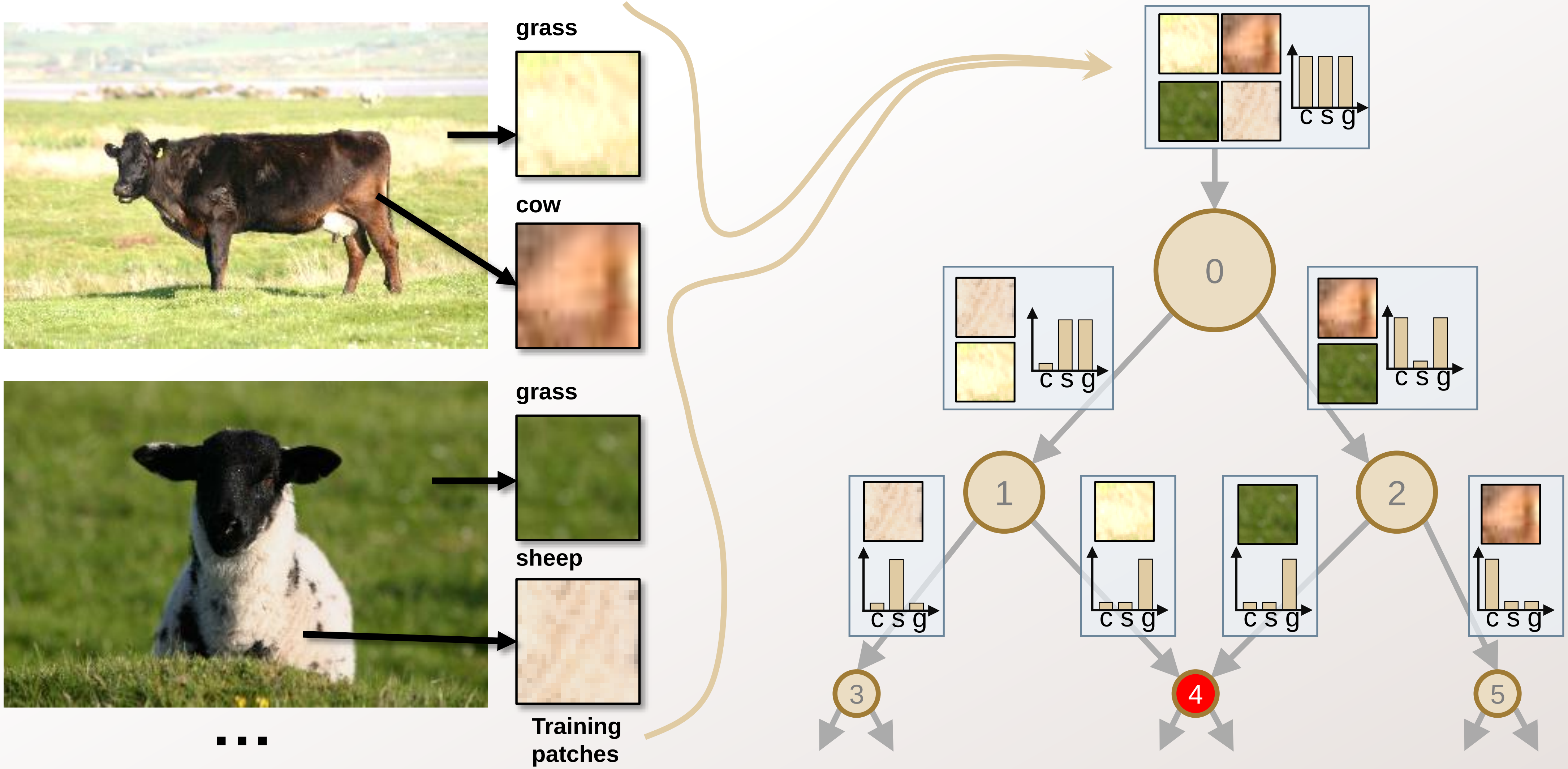


A toy example on classifying images of cows, sheep and grass



A toy example on classifying images of cows, sheep and grass

Trees vs DAGs



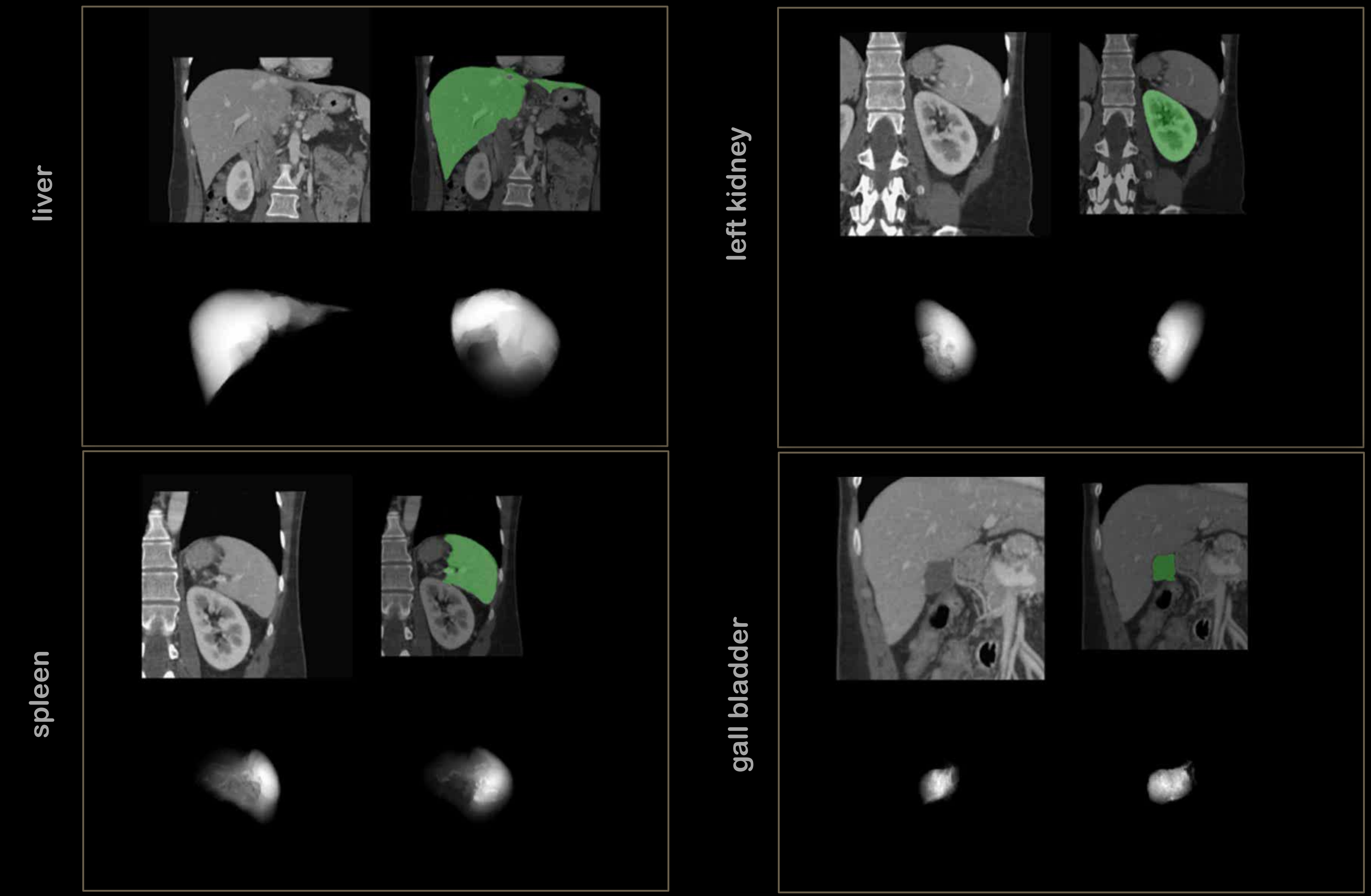
Merged nodes help capture appearance invariance

Anatomy Localization in 3D Computed Tomography Scans

Input CT scan

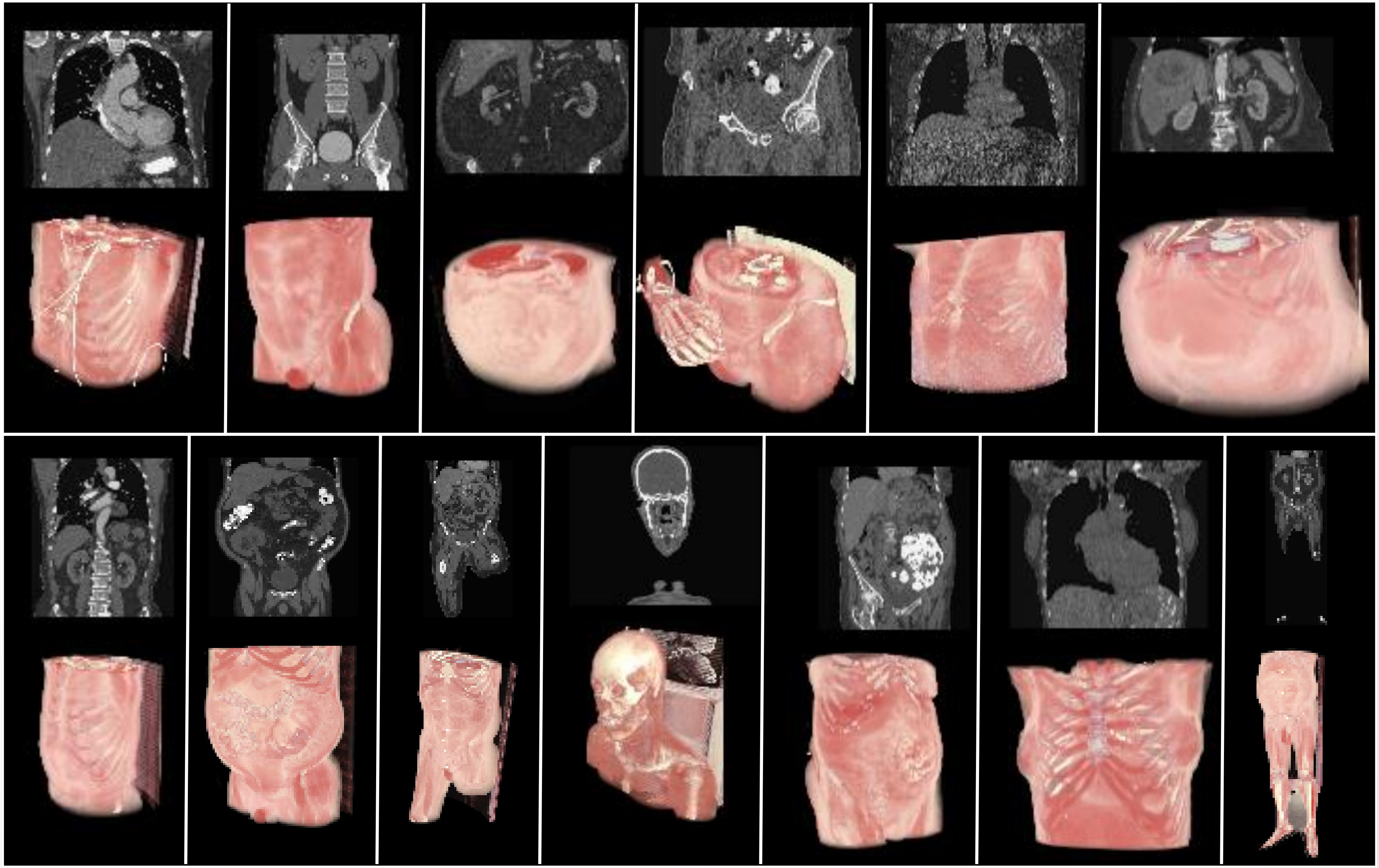
Output anatomy localization

Anatomy localization: why is it hard?



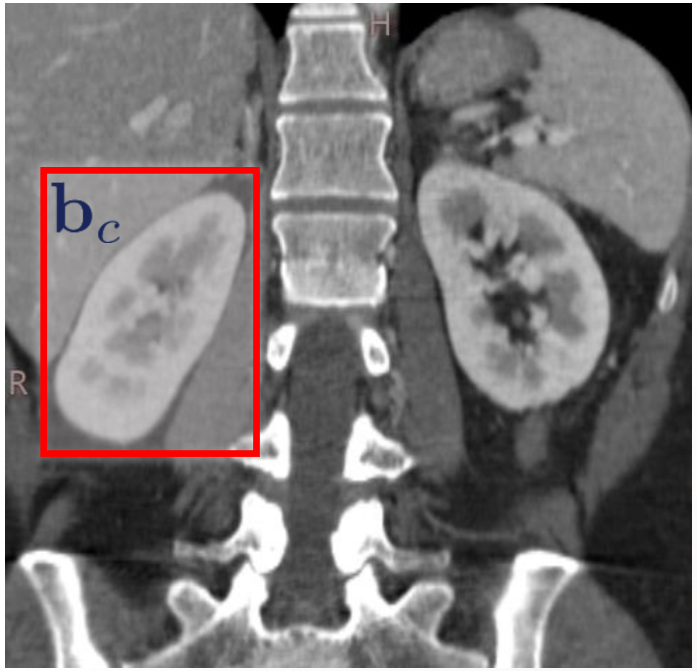
High variability in appearance, shape, location, resolution, noise, pathologies …

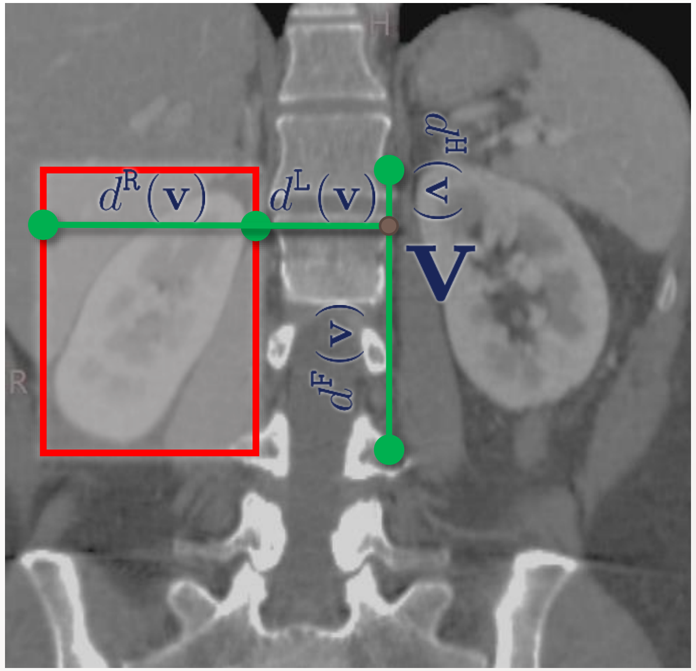
Anatomy localization: the ground-truth database

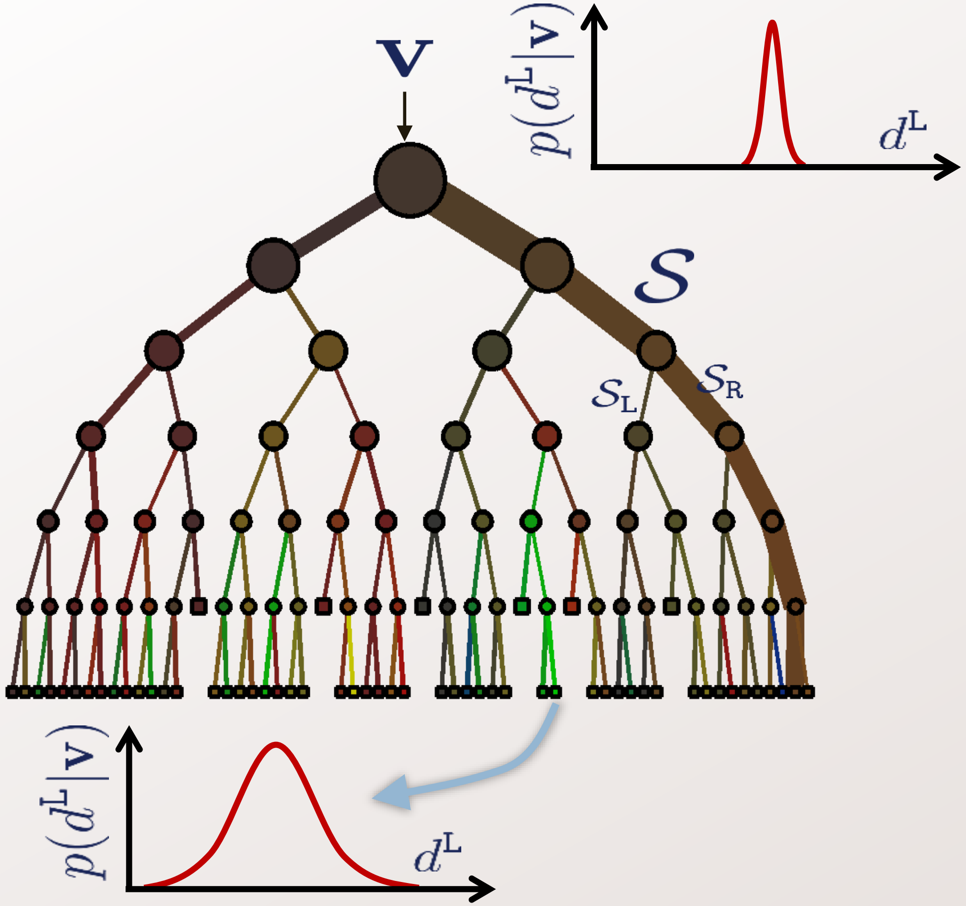


Different image cropping, noise, contrast/no-contrast, resolution, scanners, body shapes/sizes, patient position…

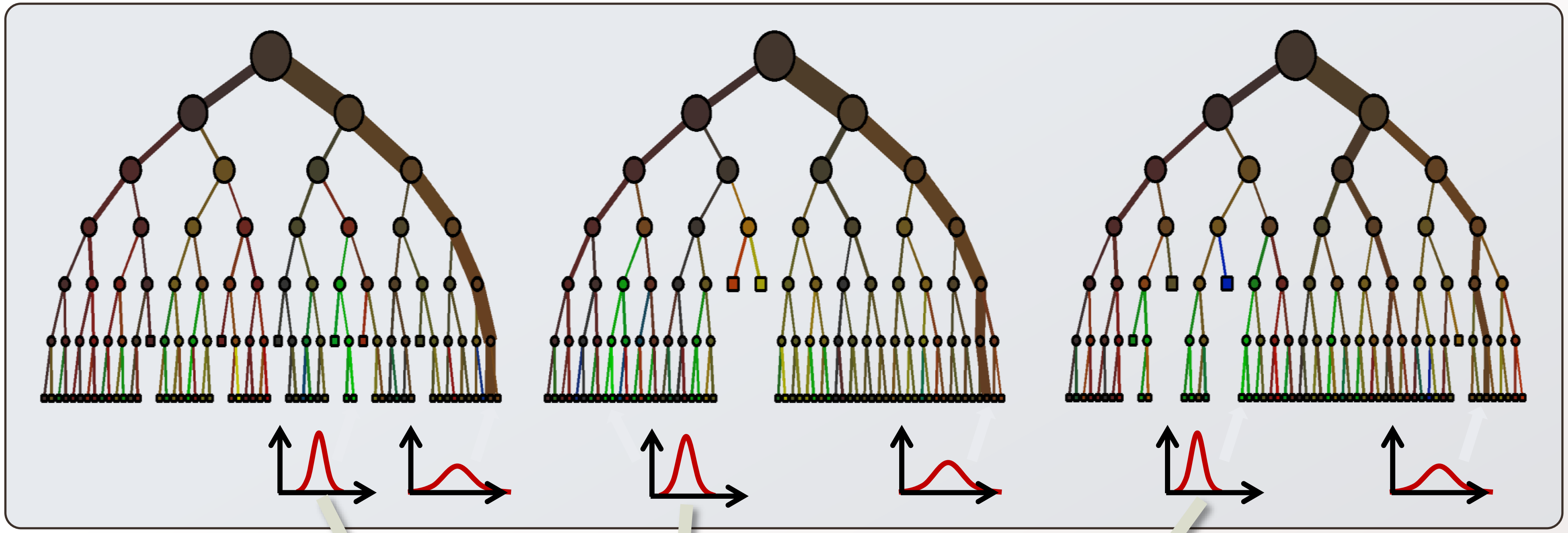
Anatomy localization: regression forest

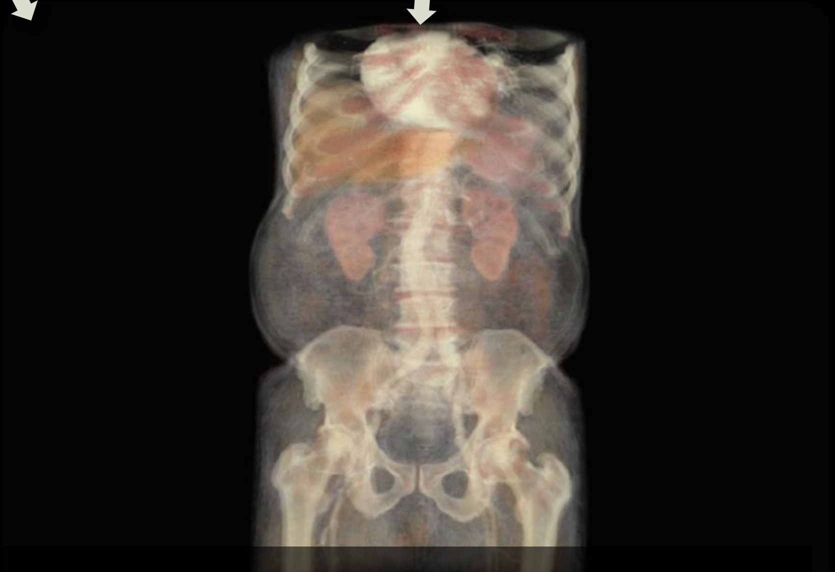






Anatomy localization: automatic landmark discovery



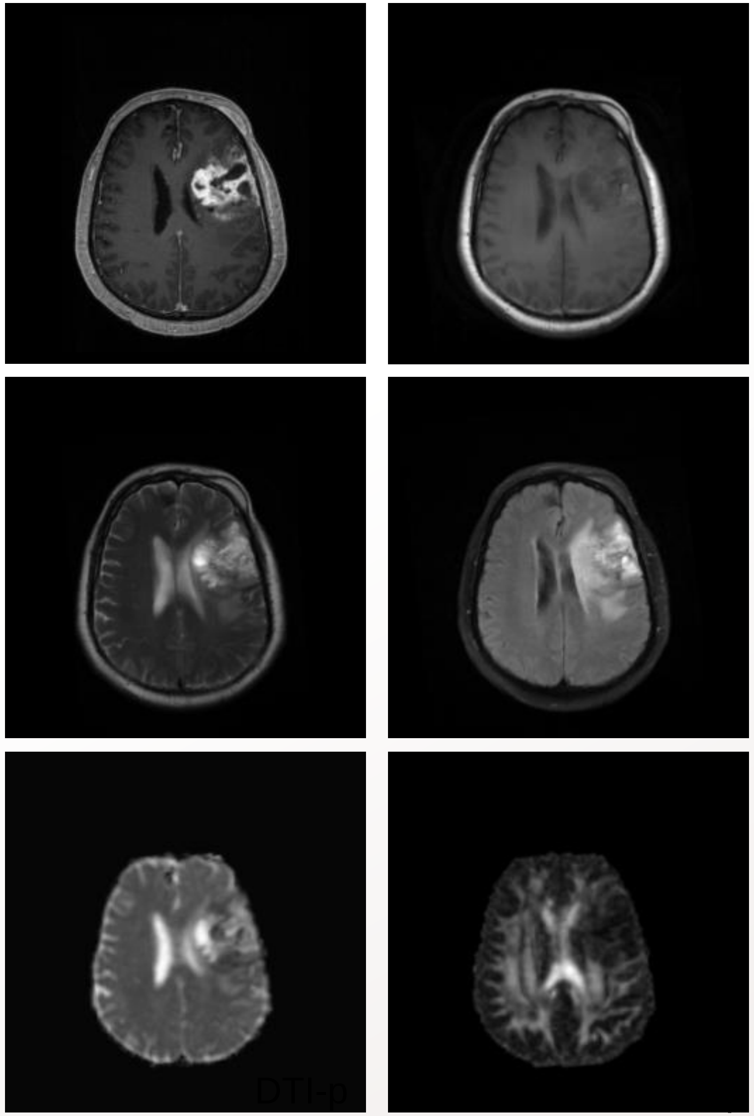


Input CT scan and detected landmark regions

Here the system is trained to detect left and right kidneys.

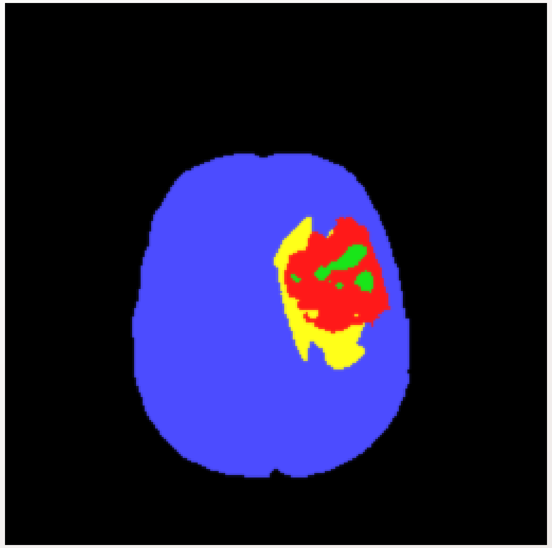
The system learns to use bottom of lung and top of pelvis to localize kidneys with highest confidence.

Automatic segmentation of brain tumour



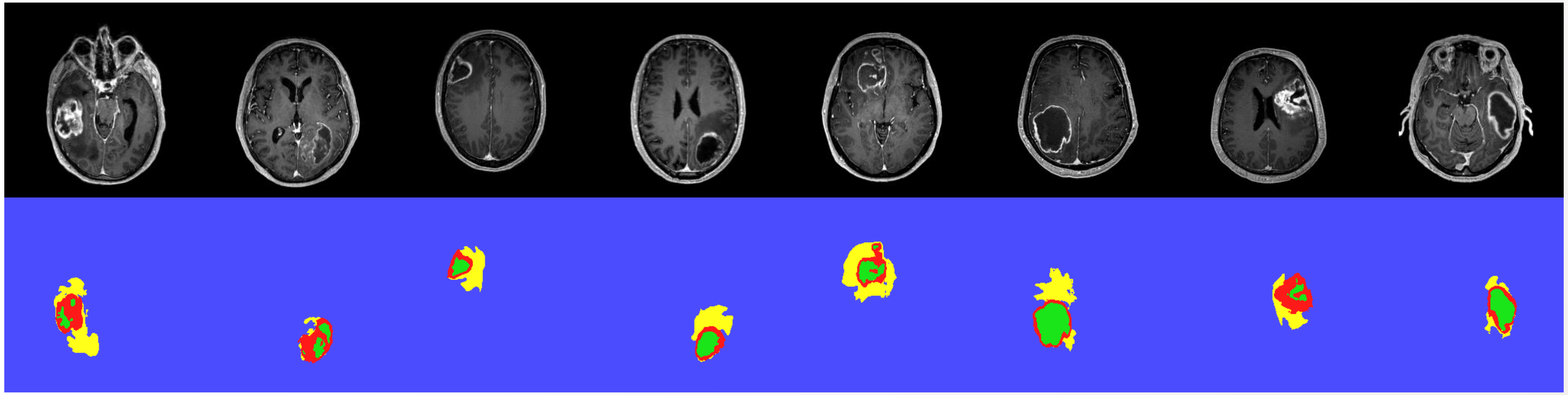
3D MRI input data

Segmentation of tumorous tissues:



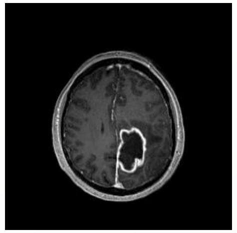
---- Active cells
---- Necrotic core
---- Edema
---- Background

Training a voxel-wise forest classifier



Tumour
Tissue
Classification

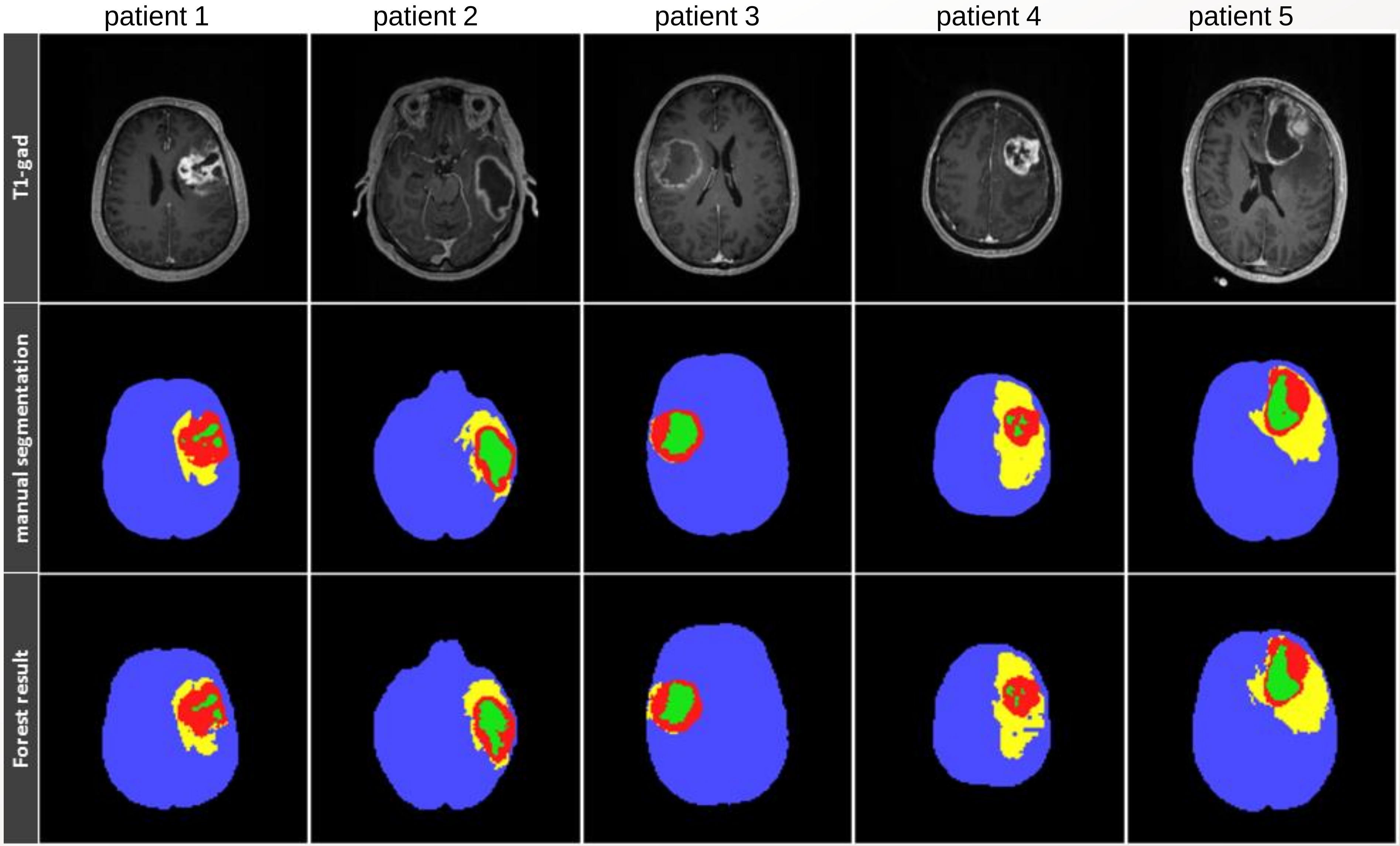
Testing the voxel-wise forest classifier



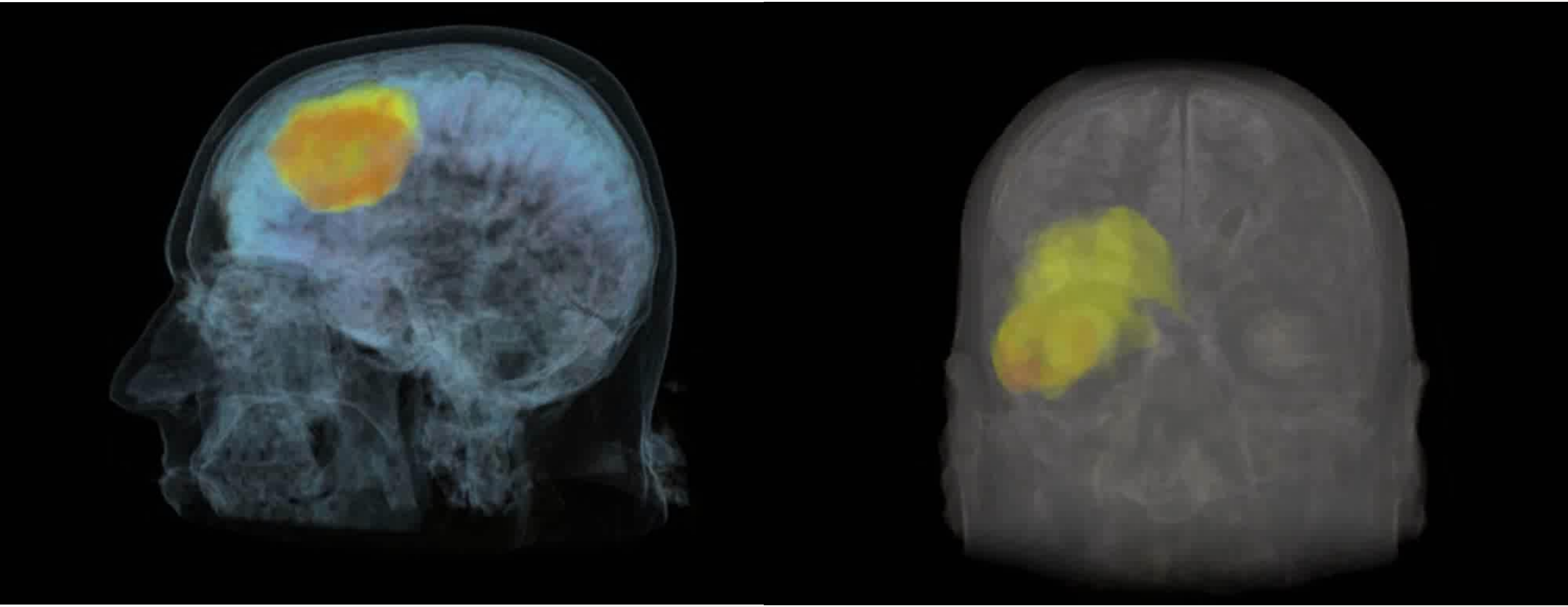
New Patient,
previously unseen

Tumour Tissue Classification

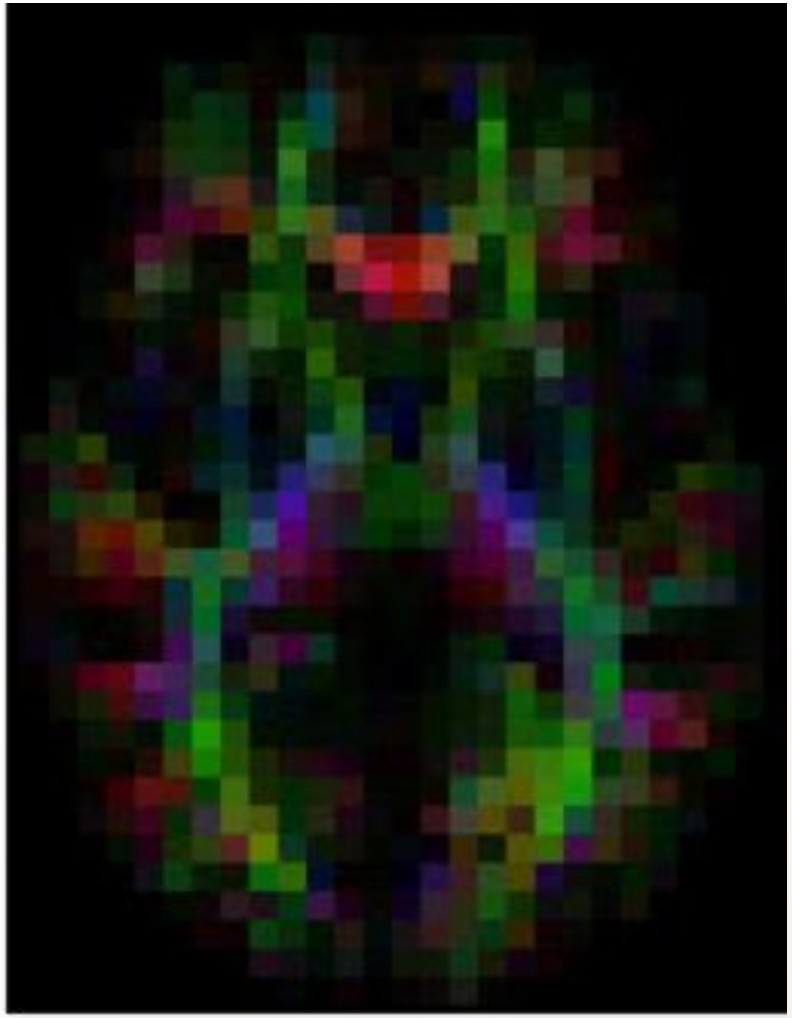
Glioblastoma segmentation results



Glioblastoma segmentation results

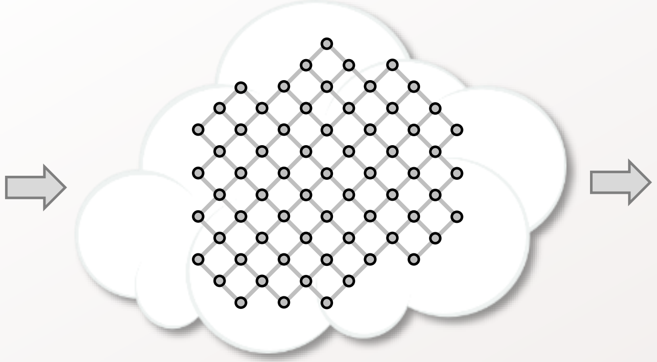


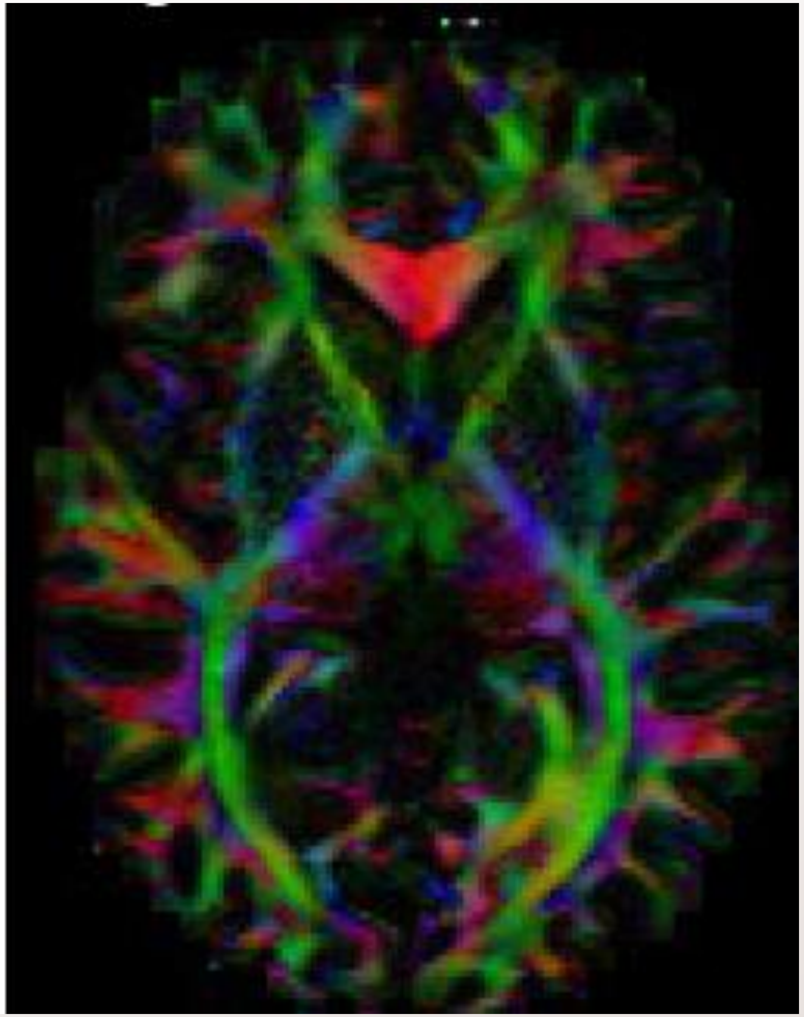
Speeding up acquisition of medical scans



Low-res diffusion MRI
(faster acquisition, cheaper)

Learned voxel predictor





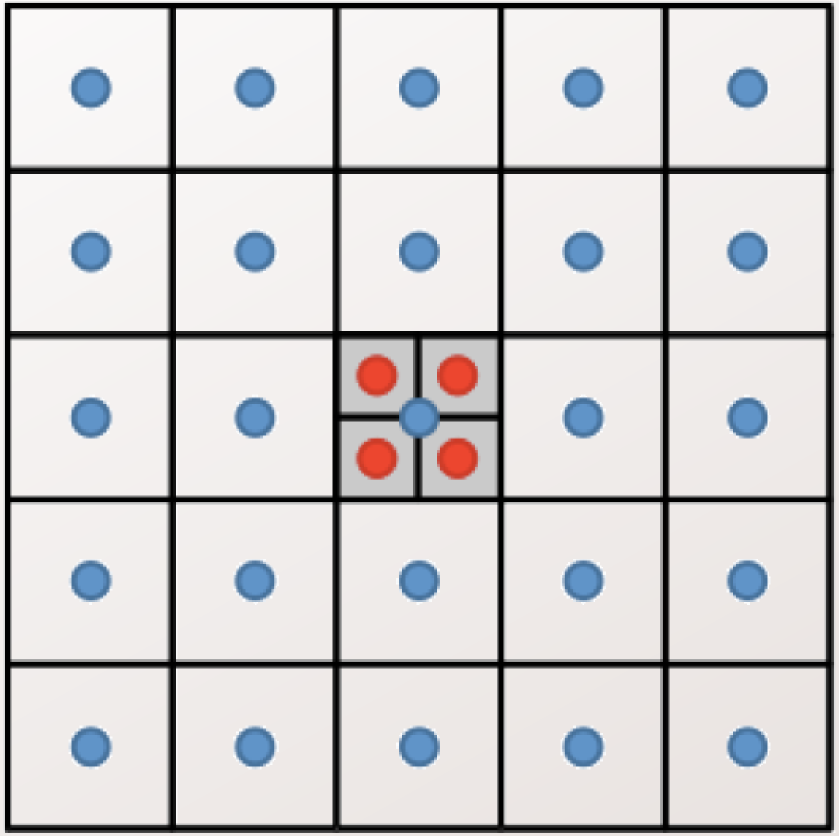
High-res diffusion MRI

Speeding up acquisition of medical scans

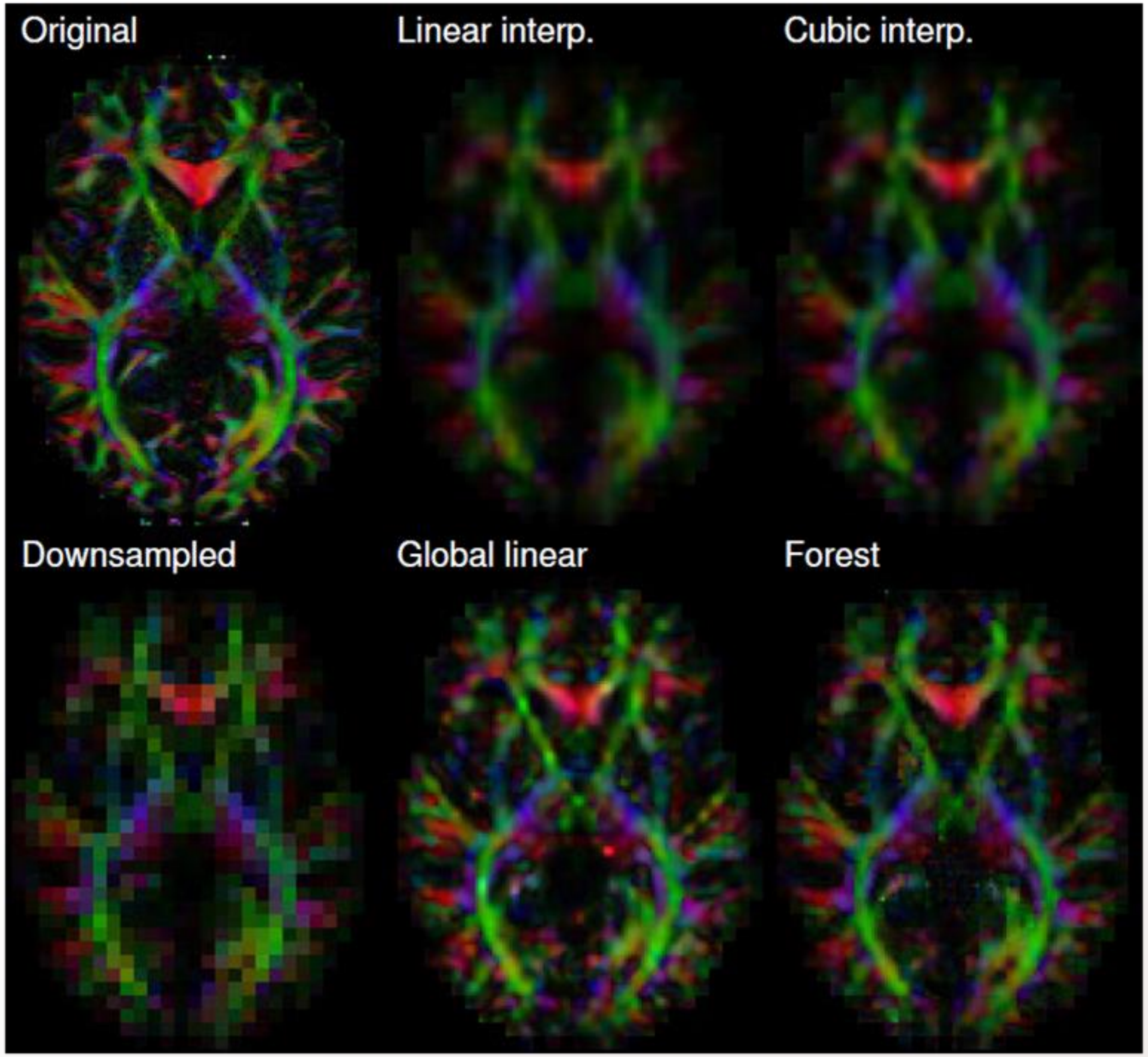
Problem statement

learning to predict the value of the **high-res voxels** from the **low-res voxels.**

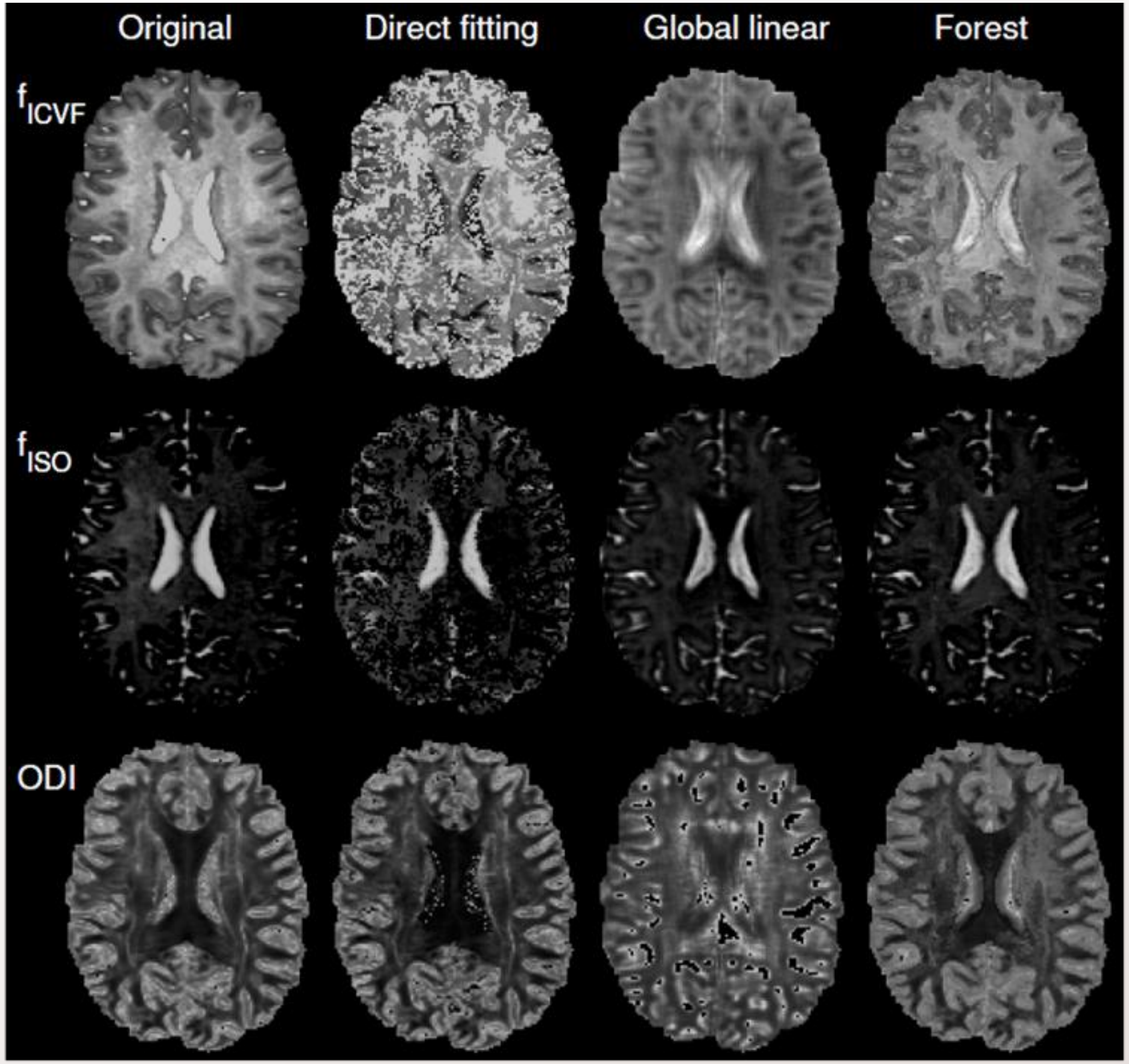
- Training data can be easily obtained
- Well defined accuracy measure



Speeding up acquisition of medical scans

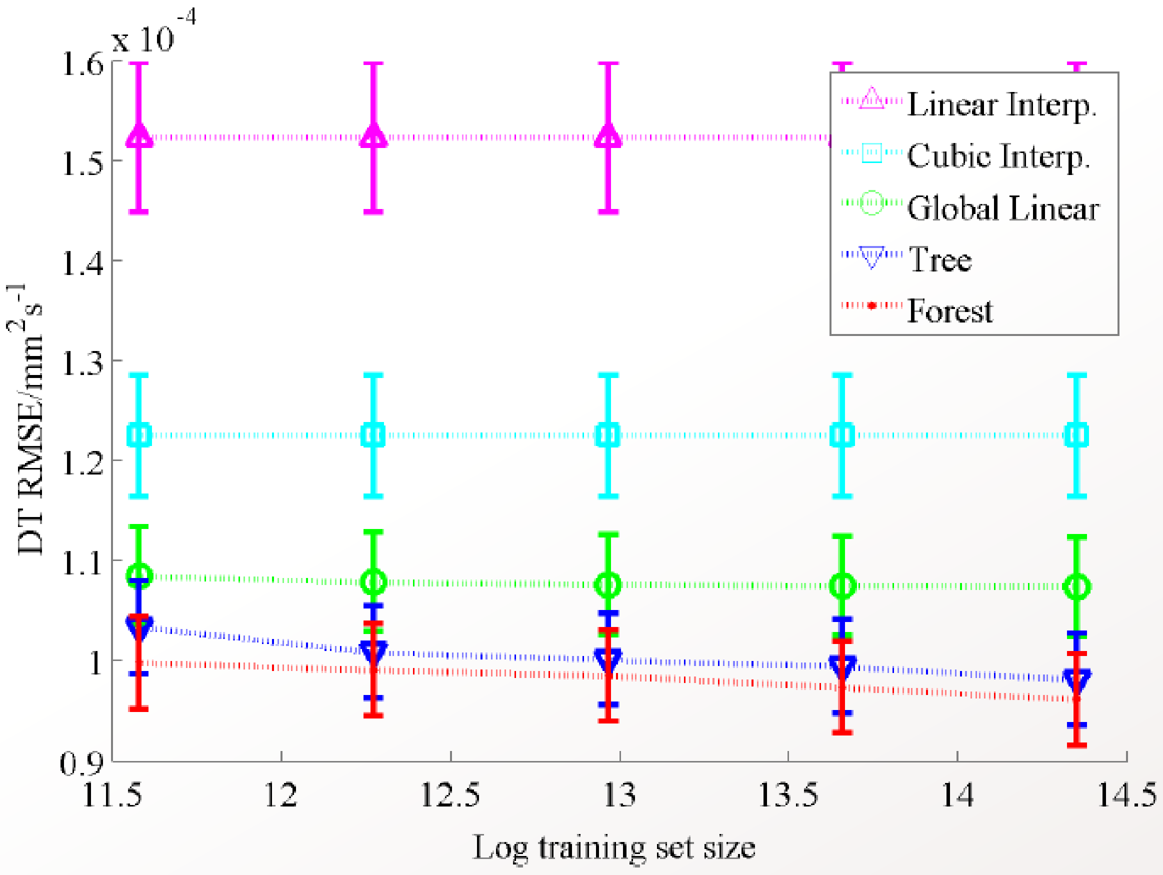


Direction-encoded colour FA maps for various reconstructed DTIs

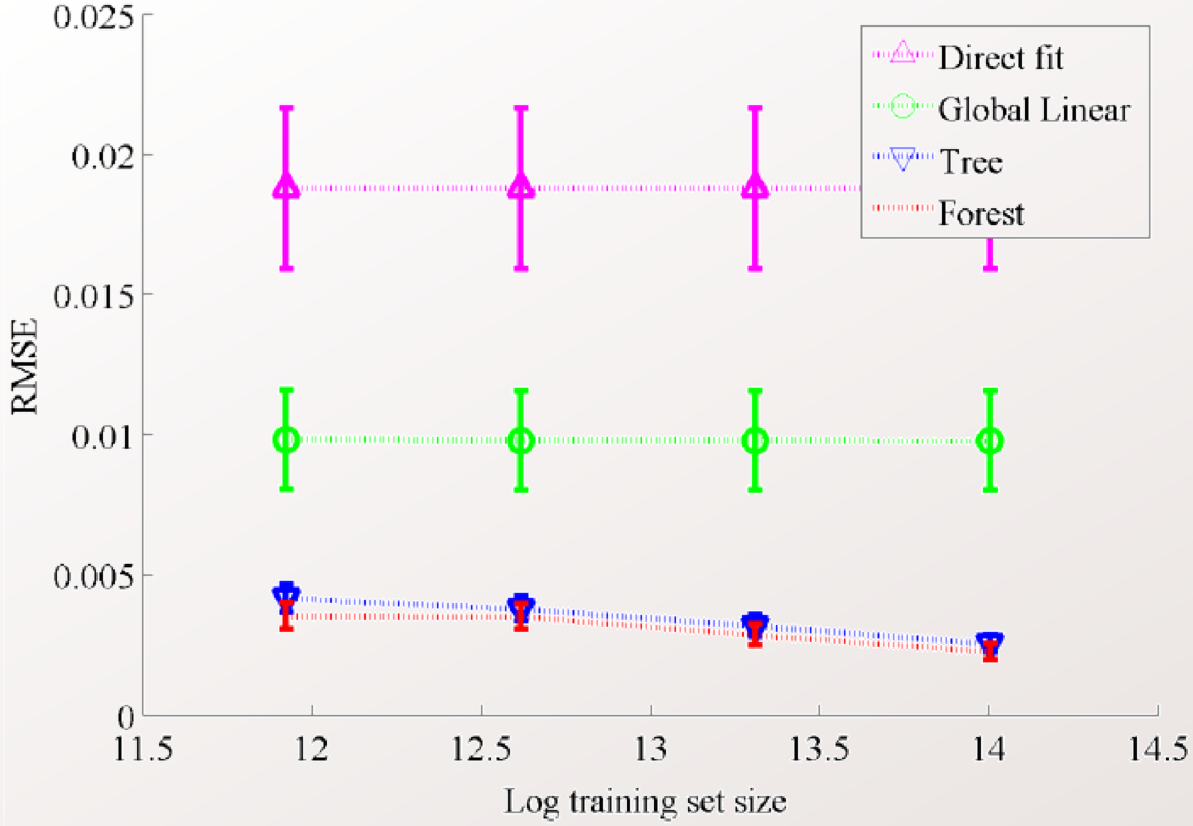


Comparison of ground truth NODDI parameter maps with various fitting techniques

Speeding up acquisition of medical scans



Reconstruction errors for DT maps



Reconstruction errors for NODDI parameter maps

D. Alexander, D. Zikic, J. Zhang, H. Zhang, and A. Criminisi, **Image Quality Transfer via Random Forest Regression: Applications in Diffusion MRI**, in *MICCAI*, Springer, 2014

Modern, efficient machine learning has the potential to revolutionize medicine!