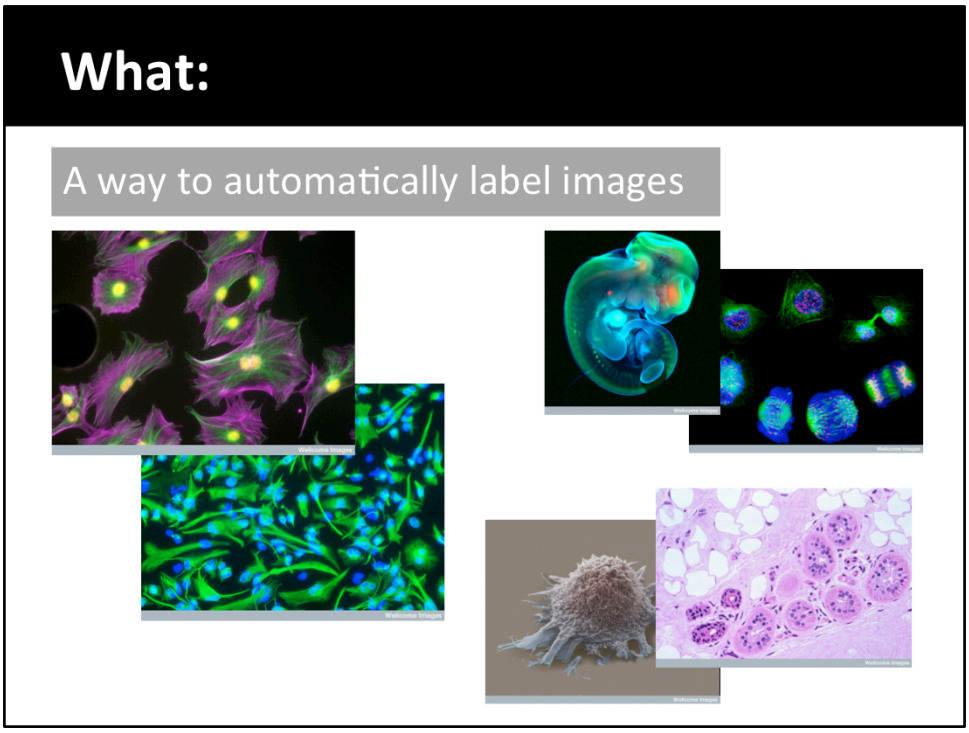
Build your own classifier with WND-CHRM and OMERO

Simon Li

Outline

- Automated classification: What and why
- How does this differ from OMERO.searcher?
- WND-CHRM: How does it work?
- Live demo
- Future development



A way of automatically annotating images within a limited domain, such as categorical tags- could be biological, experimental, image modality, etc.
E.g. Automatic tagging of metadata, automated cancer grading, looking for abnormal samples.
Some similarity to OMERO.searcher in the sense of labelling patterns, but could also be used for other tasks such as automated cancer grading.
Situations where we have a lot of images and want a way to group them.

Why?

Too much data, not enough people

More reliable
- Consistent
- Unbiased

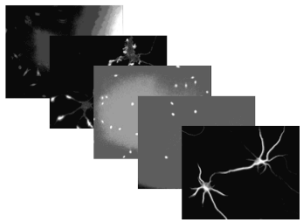
High throughput screens mean we've got more data than we can handle. Large number of samples, large number of dimensions.
Also saves time if we can make use of automated analysis.
Should be more reliability and consistent, ideally unbiased though in practice that depends on how it's been trained.
Better accuracy means subtle phenotypes can be detected. A way to explore data- if there are discernible differences between groups maybe it's worth investigating further.
Not a complete substitute for human analysts.

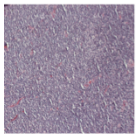
Search / classification

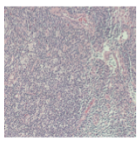
OMERO.searcher:

Database-wide image search

"Just works"







WND-CHRM:

Designed to build a classifier targeted at a *particular* problem

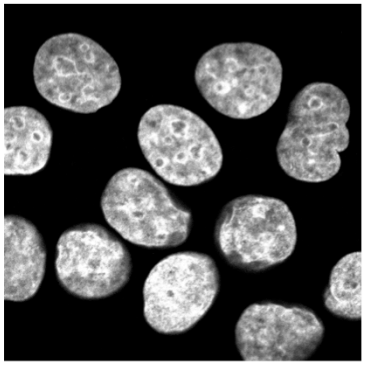
Requires significant user input

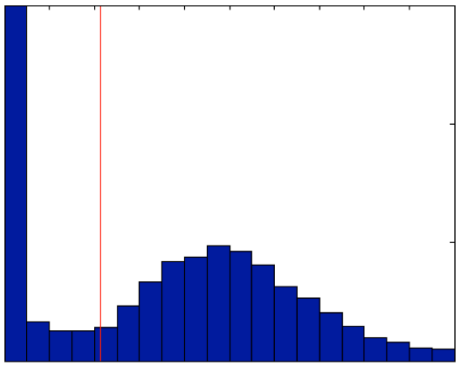
OMERO.searcher is targeted at general database-wide image search subject to the domain of the image feature set, i.e. searching a whole database for a particular class of image based on a small number of examples. It is based on the idea of pairwise image similarity and does not require image labeling (though this is useful for testing), and is dynamically updated with new data. Furthermore it should ideally be usable by someone with no technical knowledge, i.e. it "just works".
In contrast WND-CHRM is designed to be used to solve a particular well-posed biological image classification problem, such as distinguishing between different types of cancer. It therefore requires a training process involving significant user input, primarily to provide labelled example images of each class/type. WND-CHRM learns from these training images, and the resulting classifier can be used to make predictions on new images.

WND-CHRM (Ilya Goldberg, NIH)

Uses a wide range of image features

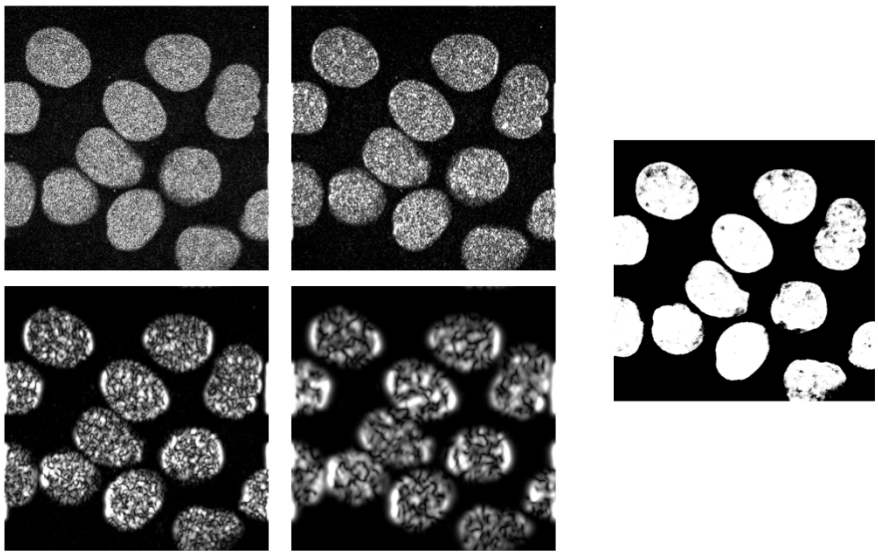
Not just biological features



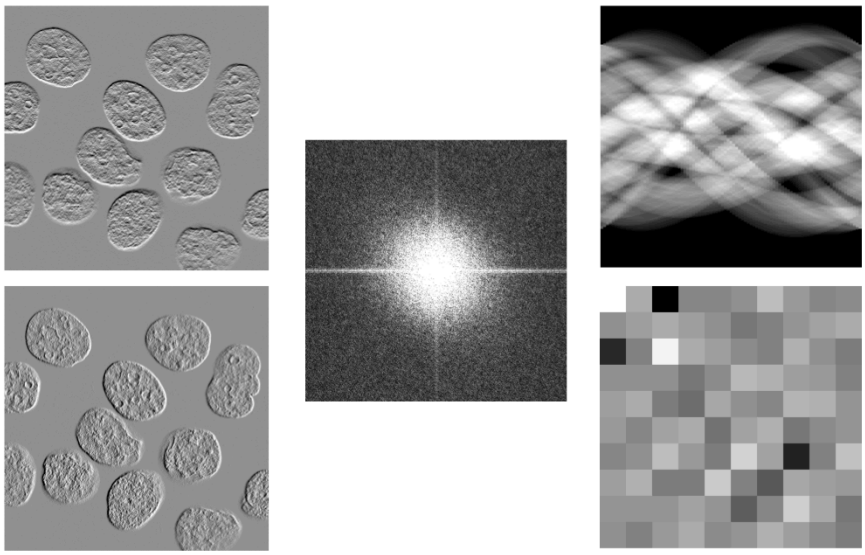


WND-CHRM
Image features: these include standard features (mean intensity, variance, several texture descriptors), and also more complex features which should be uncorrelated

WND-CHRM



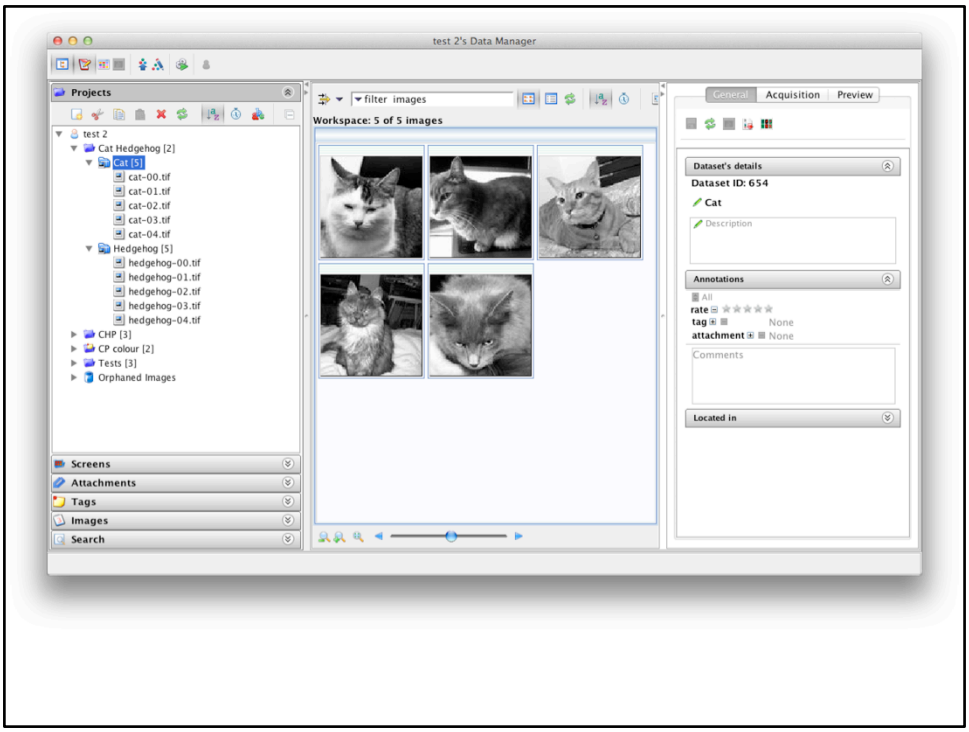
WND-CHRM

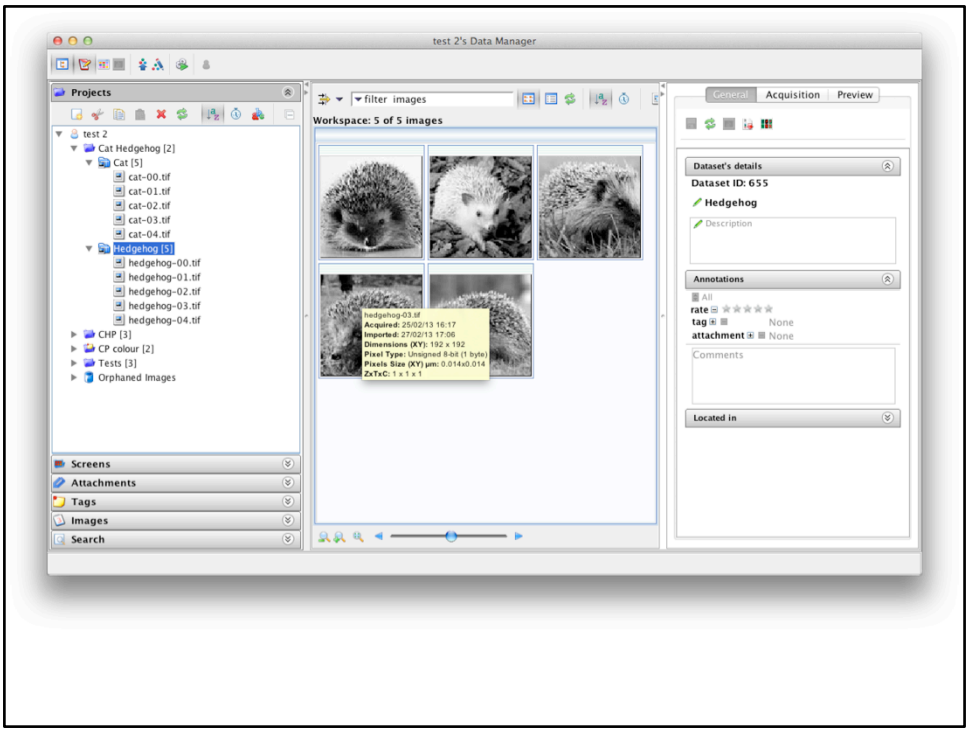


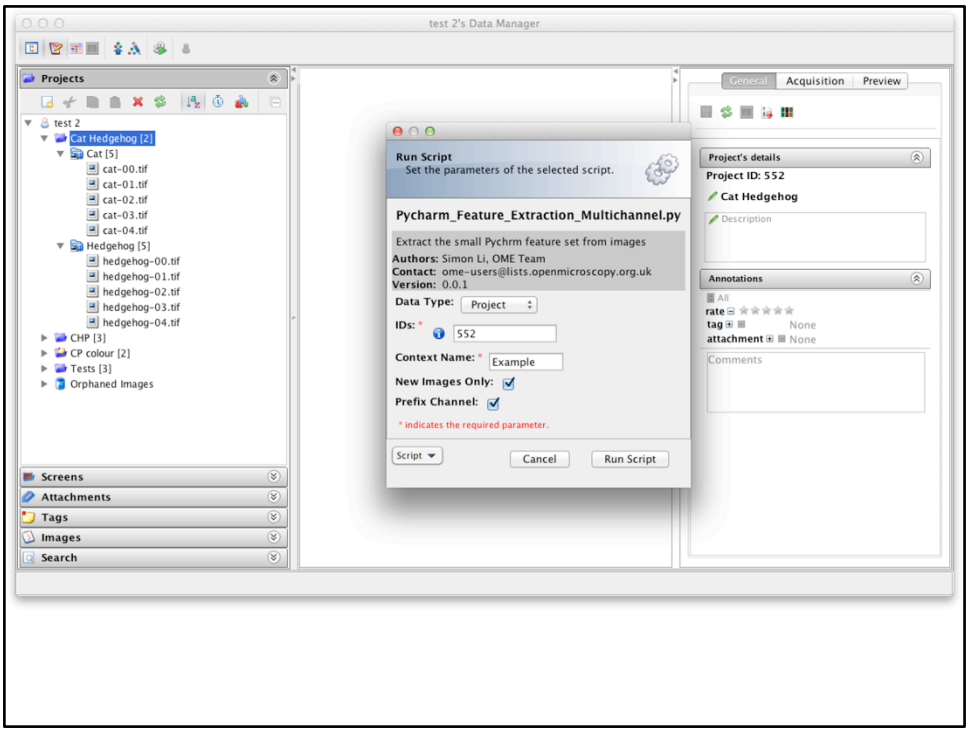
Live demo: Train

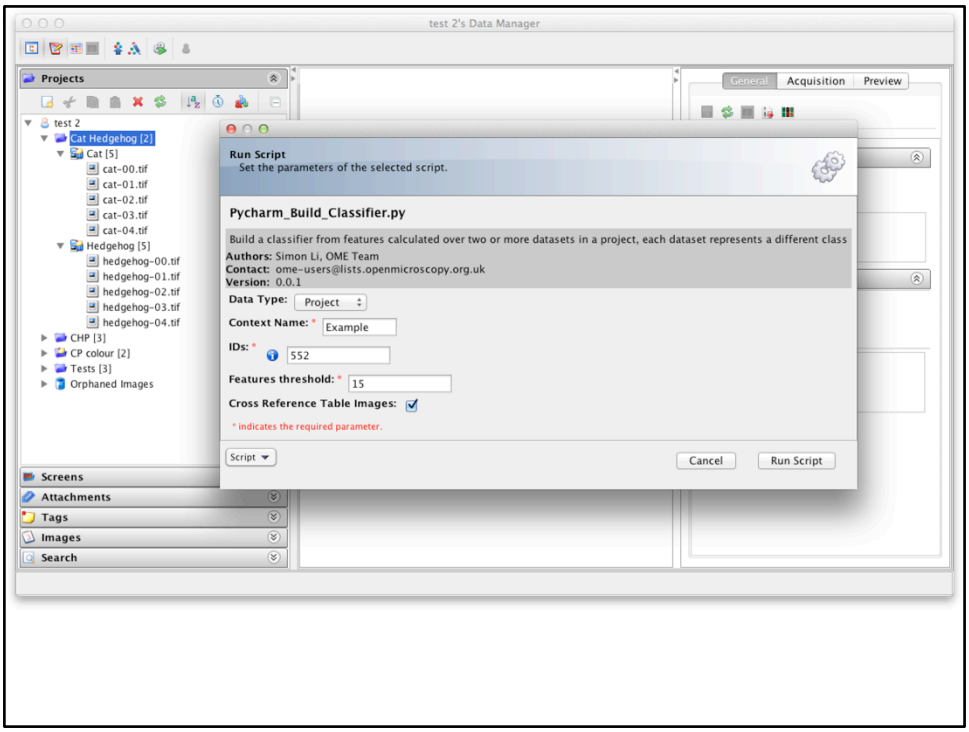
1. Create training sets
2. Calculate features
3. Train classifier

1. Arrange the training images into separate Datasets, one per category, all within a single Project
2. Select the Project, run the Pychrm_Extract_Features_Multichannel.py script. This will create a HDF5 FileAnnotation on each Dataset containing the calculated features for the images in that Dataset
3. Select the Project, run the Pychrm_Build_Classifier.py script. This will create three HDF5 FileAnnotations on the Project which hold the classifier parameters. It will also create a Tag-set named after the Project from Step 1, containing with one tag for each class (corresponding to the names of the Datasets in Step 1).


test 2's Data Manager
Projects
test 2
Cat Hedgehog [2]
Cat [5]
cat-00.tif
cat-01.tif
cat-02.tif
cat-03.tif
cat-04.tif
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Orphaned Images
Screens
Attachments
Tags
Images
Search
filter images
Workspace: 5 of 5 images
General
Acquisition
Preview
Dataset's details
Dataset ID: 654
Cat
Description
Annotations
All
rate
tag
None
attachment
None
Comments
Located in


test 2's Data Manager
Projects
test 2
Cat Hedgehog [2]
Cat [5]
cat-00.tif
cat-01.tif
cat-02.tif
cat-03.tif
cat-04.tif
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Orphaned Images
Screens
Attachments
Tags
Images
Search
filter images
Workspace: 5 of 5 images
General
Acquisition
Preview
Dataset's details
Dataset ID: 655
Hedgehog
Description
Annotations
rate
tag
None
attachment
None
Comments
Located in

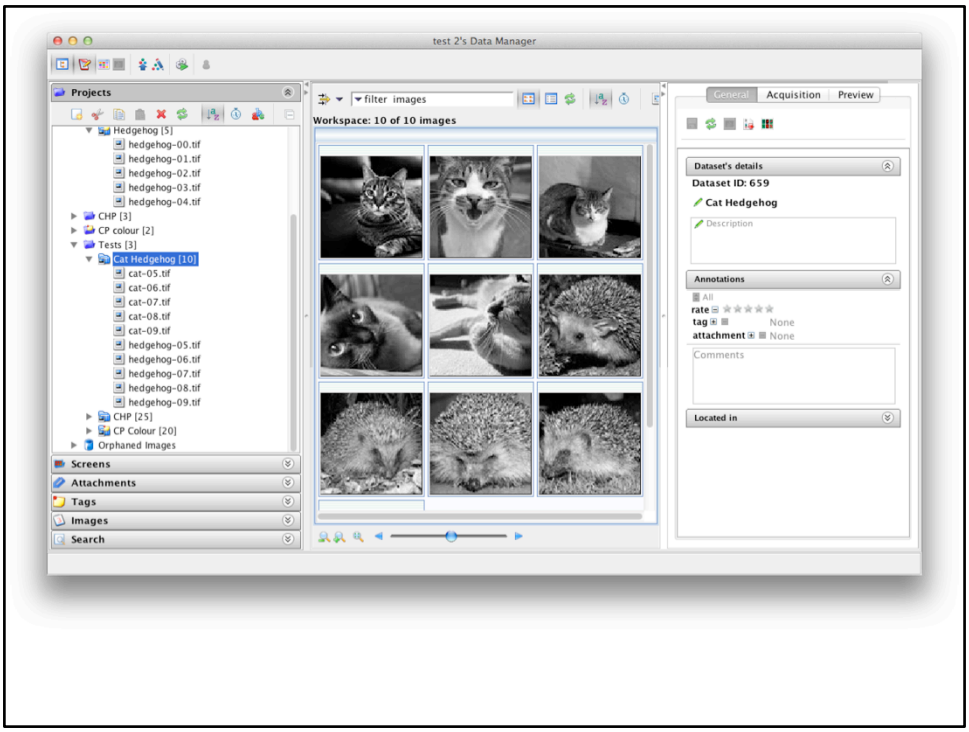

test 2's Data Manager
Projects
test 2
Cat Hedgehog [2]
Cat [5]
cat-00.tif
cat-01.tif
cat-02.tif
cat-03.tif
cat-04.tif
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Orphaned Images
Screens
Attachments
Tags
Images
Search
Run Script
Set the parameters of the selected script.
Pycharm_Feature_Extraction_Multichannel.py
Extract the small Pychrm feature set from images
Authors: Simon Li, OME Team
Contact: ome-users@lists.openmicroscopy.org.uk
Version: 0.0.1
Data Type: Project
IDs: * 552
Context Name: * Example
New Images Only:
Prefix Channel:
* indicates the required parameter.
Script
Cancel
Run Script
General
Acquisition
Preview
Project's details
Project ID: 552
Cat Hedgehog
Description
Annotations
All
rate
tag None
attachment None
Comments

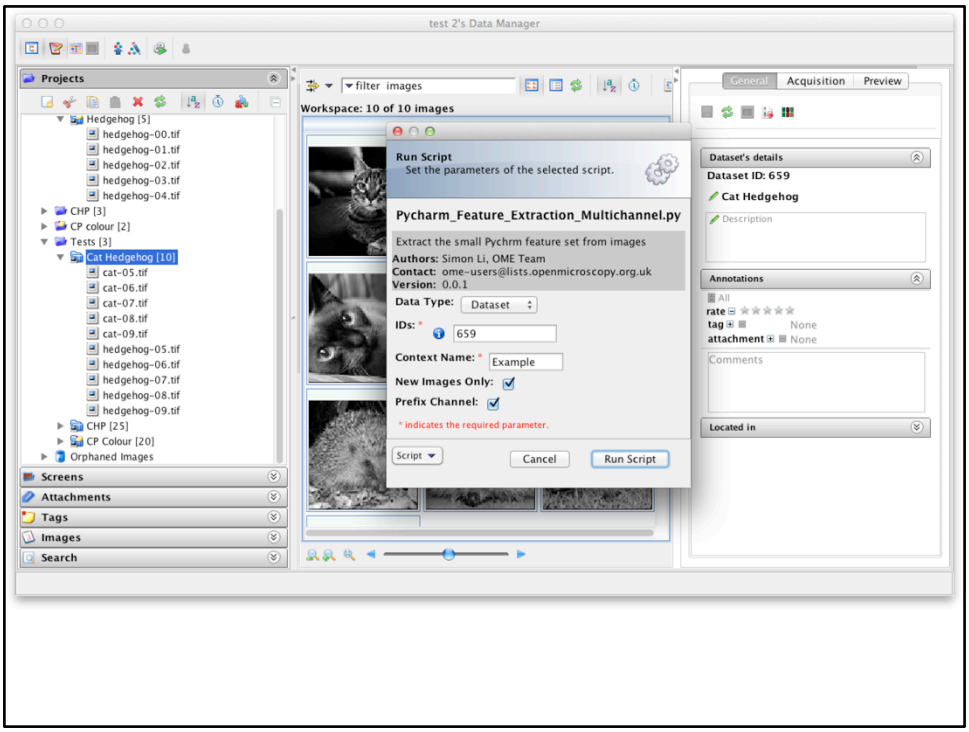

test 2's Data Manager
Projects
test 2
Cat Hedgehog [2]
Cat [5]
cat-00.tif
cat-01.tif
cat-02.tif
cat-03.tif
cat-04.tif
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Orphaned Images
Screens
Attachments
Tags
Images
Search
General
Acquisition
Preview
Run Script
Set the parameters of the selected script.
Pycharm_Build_Classifier.py
Build a classifier from features calculated over two or more datasets in a project, each dataset represents a different class
Authors: Simon Li, OME Team
Contact: ome-users@lists.openmicroscopy.org.uk
Version: 0.0.1
Data Type: Project
Context Name: * Example
IDs: * 552
Features threshold: * 15
Cross Reference Table Images:
* indicates the required parameter.
Script
Cancel
Run Script

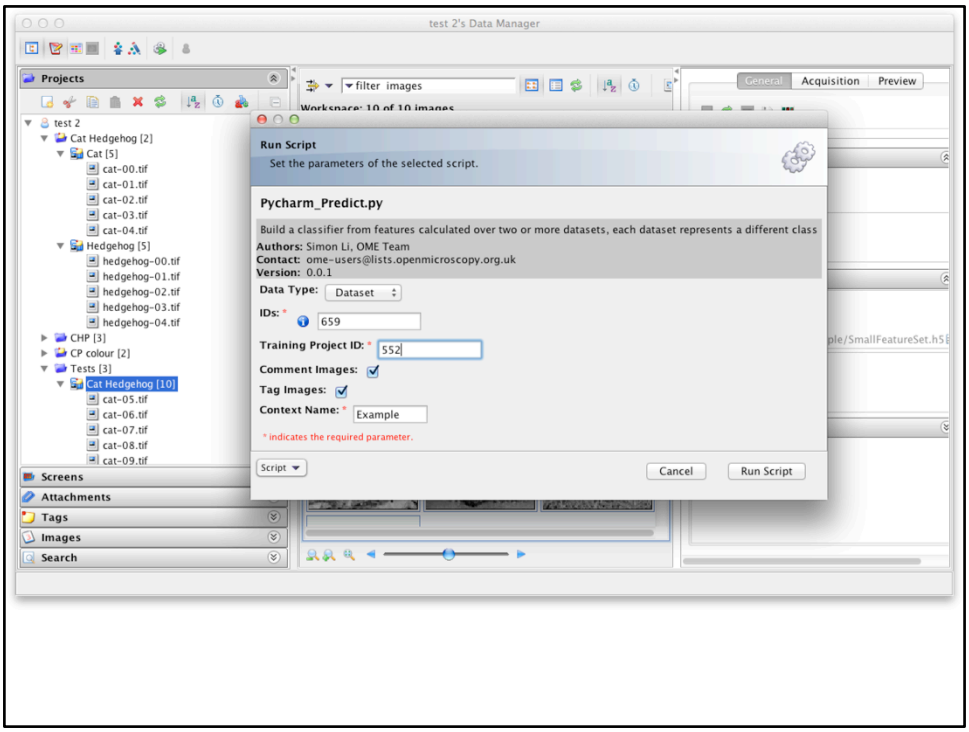
Live demo: Predict

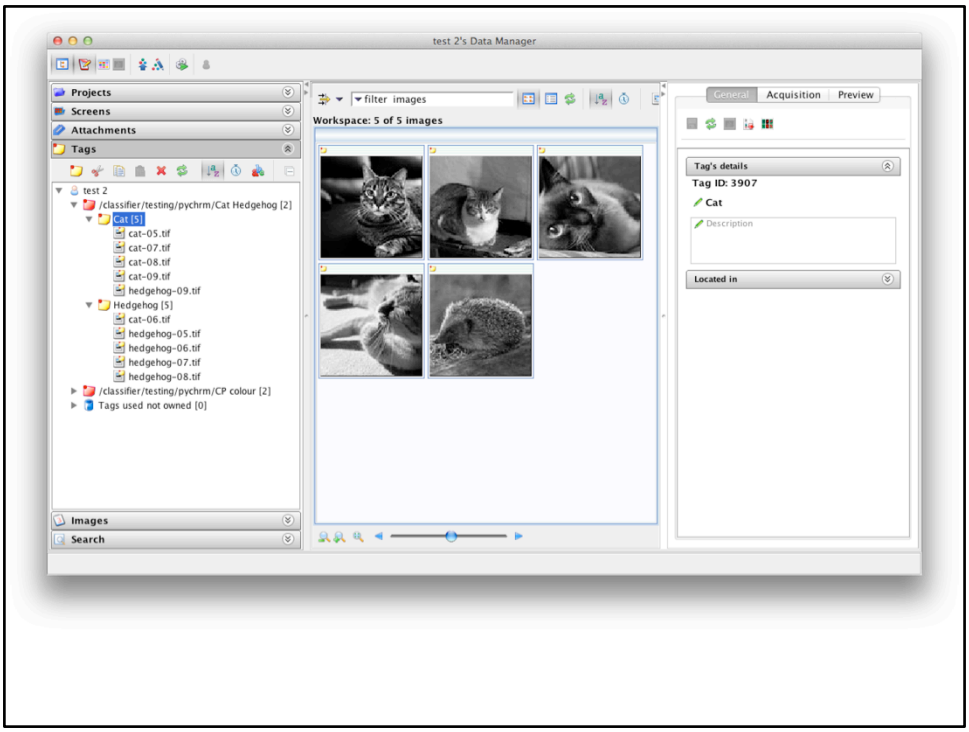
1. Select datasets to be labelled
2. Calculate features
3. Predict
4. View results

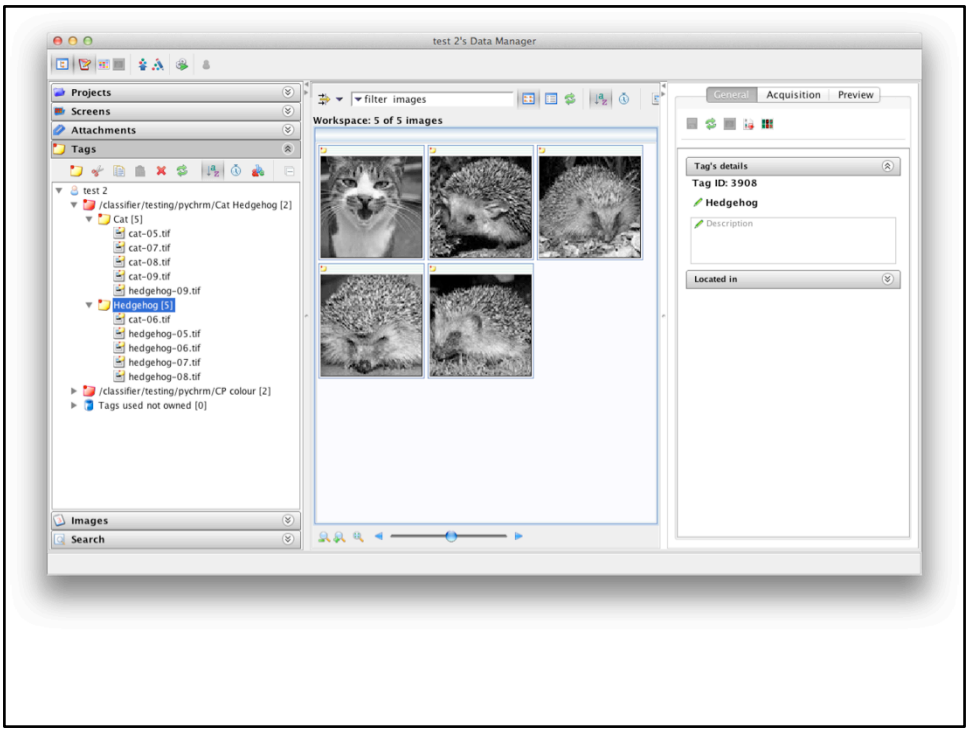
1. Run the Pychrm_Extract_Feature_Multichannel.py script on the Dataset(s) or Project(s) containing the images to be labelled
2. Select the Dataset(s) or Project(s) to be labelled, run the Pychrm_Predict.py script manually entering the Project-ID of the training project
3. By default the classifier will tag images with the predicted label, and also add a comment containing the probabilities of each class.
4. Predictions can also be browsed from the Tags view in the Insight or Web clients.


test 2's Data Manager
Projects
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Cat Hedgehog [10]
cat-05.tif
cat-06.tif
cat-07.tif
cat-08.tif
cat-09.tif
hedgehog-05.tif
hedgehog-06.tif
hedgehog-07.tif
hedgehog-08.tif
hedgehog-09.tif
CHP [25]
CP Colour [20]
Orphaned Images
Screens
Attachments
Tags
Images
Search
filter images
Workspace: 10 of 10 images
General
Acquisition
Preview
Dataset's details
Dataset ID: 659
Cat Hedgehog
Description
Annotations
All
rate
tag
None
attachment
None
Comments
Located in


test 2's Data Manager
Projects
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Cat Hedgehog [10]
cat-05.tif
cat-06.tif
cat-07.tif
cat-08.tif
cat-09.tif
hedgehog-05.tif
hedgehog-06.tif
hedgehog-07.tif
hedgehog-08.tif
hedgehog-09.tif
CHP [25]
CP Colour [20]
Orphaned Images
Screens
Attachments
Tags
Images
Search
filter images
Workspace: 10 of 10 images
Run Script
Set the parameters of the selected script.
Pycharm_Feature_Extraction_Multichannel.py
Extract the small Pychrm feature set from images
Authors: Simon Li, OME Team
Contact: ome-users@lists.openmicroscopy.org.uk
Version: 0.0.1
Data Type: Dataset
IDs: * 659
Context Name: * Example
New Images Only:
Prefix Channel:
* indicates the required parameter.
Script
Cancel
Run Script
General
Acquisition
Preview
Dataset's details
Dataset ID: 659
Cat Hedgehog
Description
Annotations
All
rate
tag None
attachment None
Comments
Located in


test 2's Data Manager
Projects
test 2
Cat Hedgehog [2]
Cat [5]
cat-00.tif
cat-01.tif
cat-02.tif
cat-03.tif
cat-04.tif
Hedgehog [5]
hedgehog-00.tif
hedgehog-01.tif
hedgehog-02.tif
hedgehog-03.tif
hedgehog-04.tif
CHP [3]
CP colour [2]
Tests [3]
Cat Hedgehog [10]
cat-05.tif
cat-06.tif
cat-07.tif
cat-08.tif
cat-09.tif
Screens
Attachments
Tags
Images
Search
filter images
General
Acquisition
Preview
Run Script
Set the parameters of the selected script.
Pycharm_Predict.py
Build a classifier from features calculated over two or more datasets, each dataset represents a different class
Authors: Simon Li, OME Team
Contact: ome-users@lists.openmicroscopy.org.uk
Version: 0.0.1
Data Type: Dataset
IDs: * 659
Training Project ID: * 552
Comment Images:
Tag Images:
Context Name: * Example
* indicates the required parameter.
Script
Cancel
Run Script


test 2's Data Manager
Projects
Screens
Attachments
Tags
test 2
/classifier/testing/pychrm/Cat Hedgehog [2]
Cat [5]
cat-05.tif
cat-07.tif
cat-08.tif
cat-09.tif
hedgehog-09.tif
Hedgehog [5]
cat-06.tif
hedgehog-05.tif
hedgehog-06.tif
hedgehog-07.tif
hedgehog-08.tif
/classifier/testing/pychrm/CP colour [2]
Tags used not owned [0]
Images
Search
filter images
Workspace: 5 of 5 images
General
Acquisition
Preview
Tag's details
Tag ID: 3907
Cat
Description
Located in


test 2's Data Manager
Projects
Screens
Attachments
Tags
test 2
/classifier/testing/pychrm/Cat Hedgehog [2]
Cat [5]
cat-05.tif
cat-07.tif
cat-08.tif
cat-09.tif
hedgehog-09.tif
Hedgehog [5]
cat-06.tif
hedgehog-05.tif
hedgehog-06.tif
hedgehog-07.tif
hedgehog-08.tif
/classifier/testing/pychrm/CP colour [2]
Tags used not owned [0]
Images
Search
filter images
Workspace: 5 of 5 images
General
Acquisition
Preview
Tag's details
Tag ID: 3908
Hedgehog
Description
Located in

Goal for OMERO 4.5

OMERO.searcher and WND-CHRM working

Intuitive user-interface

- Accessible to anyone

Aim is to have a tool which is useful for researchers, and can produce scientifically sound results.

Longer term

- Unified feature API
- Prediction of continuous variables
- Cluster support
- Display useful features
- What do users and developers want?

Unified API, so OMERO.searcher can use WND-CHRM features and vice-versa, and also other 3rd party features
Continuous predictions, e.g. expected survival time based on a cancer biopsy
Cluster support, speed up training stage, especially feature calculation, allow real-time predictions
Algorithm developers, instead of being a place to release algorithms maybe OMERO should support people who want to develop directly against OMERO, e.g. providing support for algorithm evaluation metrics such as cross-validation tools