02-261 – Quantitative Cell and Molecular Biology Lab

Programming Lab: High Content Screening

In microscopy experiments, we cannot simply rely on the appearance of images to make strong conclusions. We require computational methods to quantify the observed phenomena. In some situations, a single measurement is not sufficient to capture the variability of an observation. In this case, researchers may use machine learning methods to identify experimental outcomes. To do so, one must calculate features which are numerical descriptors of the images. Then a model can be learned to predict the labels of images based on the values of these features. In this Lab, you will design and program feature calculation methods and assess the accuracy of models learned using those features.

You have been provided with two image sets:

1. Manual HCS Images – refer to the shared data table for the source information for these images.
2. Automated HCS Images – This includes a set of images generated on an automated microscope which had the capability for widefield and confocal imaging.

These images contain NIH 3T3 cells. Some of which were treated with FCCP which is known to disrupt mitochondrial membrane potential. The cells were stained with Hoechst (DNA) and Mitotracker FM GFP.

Channel	**Microscopy Method**	**Stain/Organelle**
d0	Widefield	Hoechst/Nucleus
d1	Widefield	Mitotracker FM Green/Mitochondria
d2	Confocal	Hoechst/Nucleus
d3	Confocal	Mitotracker FM Green/Mitochondria

In order to classify these images, we need to calculate image features.

If you are in the 9 unit course or 12 unit course, you need to complete the following:

Generate a set of features designed to detect the difference between FCCP treated and untreated *fields* in the manually generated data. You should use at least four types of features.

Finally, you need to generate code for testing models learned from your features using cross-validation. You should use a Random Forest classifier.

If you are in the 12 unit course, you will need to complete these additional tasks

Design a set of features detect the difference between widefield and confocal *images* regardless of the stain.

Programming Assignment: For each classification problem, you will submit the following:

1. Feature Calculation Description – Describe the process for the calculation of features including images or figures where helpful. Explain why you chose these features.

2. Show the feature importance for each feature. Comment on importances relative to your expectations. Why do you think some features were more important than others?
3. Generate a ROC curve showing the accuracy of the model you generated using 5 fold cross-validation. In order to generate the curve, combine all predictions from all images when they are in the held out set.
4. Look at fields which were misclassified and qualitatively describe differences between those images and fields which were correctly classified. Describe whether these issues could be addressed with better imaging/laboratory technique or if you could address these issues computationally through better feature calculations.
5. Commented code.

Submission: You will submit your assignment in two parts. The first part should be a PDF. The second submission will be a python file named uniquely for you.

Potentially Useful information:

Hoechst 33342

Mitotracker Green FM

FCCP

Mitochondrial depolarization with FCCP

Matplotlib is a library designed for making nice plots.

Numpy is a library designed for math with a special emphasis on arrays.

Scipy is a library built around numpy for specialized scientific tasks.

Skimage is a library for image processing tasks.

You may or may not need to use any of these sources of information, but they may prove useful.

Loading Images with Scipy

Contrast Stretching with skimage

Image processing with Scipy

Otsu Thresholding with Scipy

Watershed Segmentation with Scipy

Gallery of matplotlib plots with Examples

More segmentation

Calculating size of objects

Edge detection

Feature Calculation