

## Machine Learning Module Operation

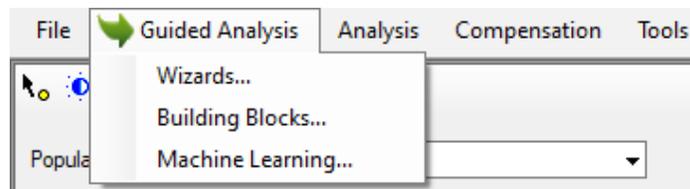
### 1. Overview

The machine learning (ML) module for IDEAS® creates an experiment specific feature that will distinguish populations based on user input. It is designed to simplify analysis by allowing users to visually create populations, and enhance discrimination by allowing users to combine multiple fluorochromes and multiple morphologies into a single super-feature.

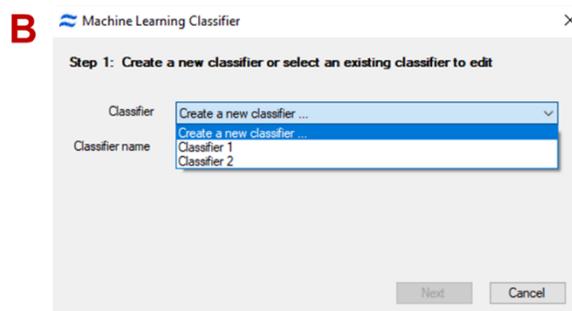
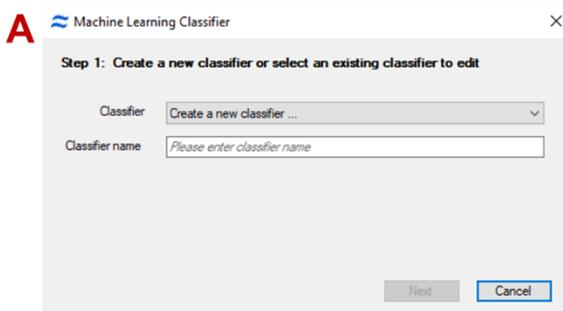
The user provides 'truth' populations, and the module creates custom 'classifier features' based on the Linear Discriminant Analysis ML. This algorithm will identify images similar to the 'truth' populations. In addition to incorporating user-defined features already present in the data file, the ML module creates a comprehensive set of additional masks and features, which it then ranks and combines with appropriate weights, to maximally separate each 'truth' population from the others. The output is a series of 'classifier features' that are specifically targeted to the desired phenotypes, and can be plotted on histograms or bivariate plots.

### 2. Calculating new features using the Machine Learning for IDEAS® module

To begin, open the data file of interest with the compensation matrix applied. You may also apply an analysis template if one exists. Once the file is open, select the "Machine Learning" option from the Guided Analysis dropdown.



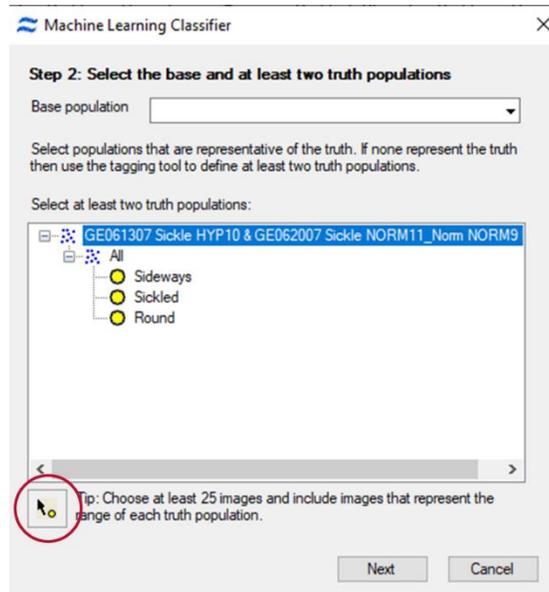
Step 1 prompts the user to either create a new classifier or modify an existing one. If no classifiers exist, one must be created (A). However, if there are existing classifiers, one may be edited by selecting it from the dropdown menu (B).





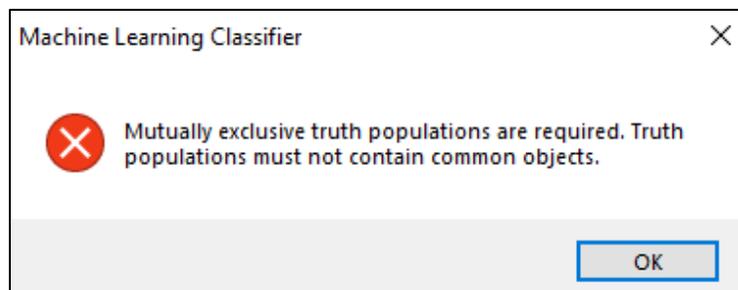
Step 2 prompts the user to select the base population from the dropdown menu at the top of the dialogue box. The “All” population will be the only option if no previous analysis template has been applied. But, if the user has applied a template where gating has been performed, other base populations can be selected (e.g. single cells).

Step 2 also prompts the user to select the populations of interest to differentiate between. If populations of interest have already been tagged, use control-click to select at least two populations. If populations of interest have not been tagged, use the tagging tool (red circle) to create them.



**Note:** At least 25 images must be tagged for the machine learning module to run. In most cases, larger numbers of images provided to the ML module will generate more robust classifier features.

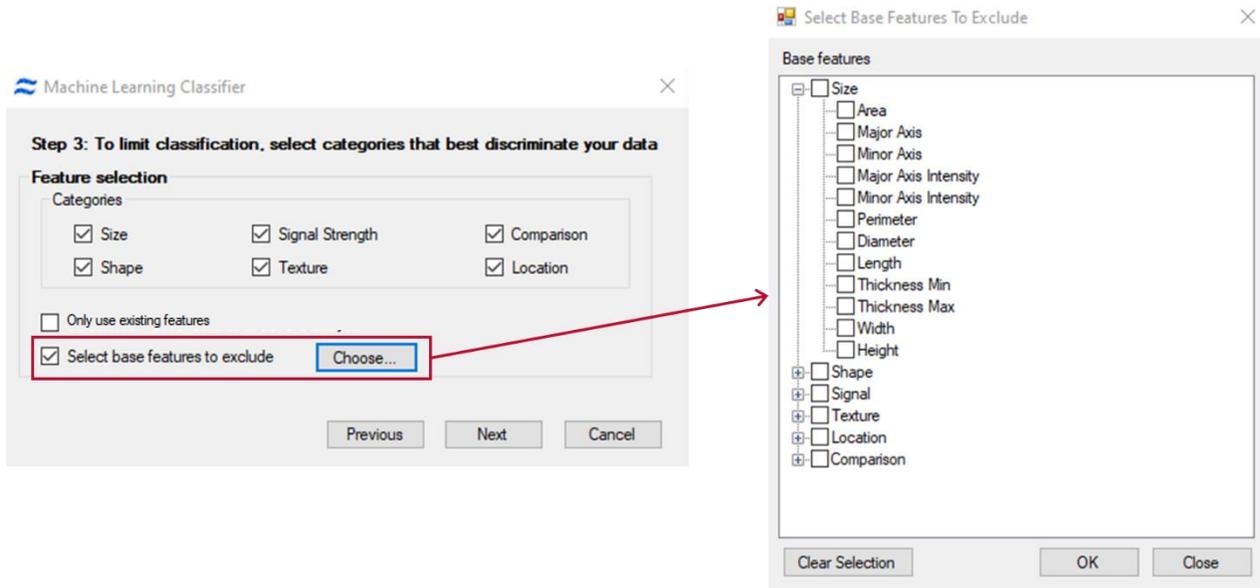
**Note:** Images in all selected populations must be mutually exclusive (i.e. one image cannot belong to two populations). The ML module will verify this before continuing and will give the following error if it detects non-mutually exclusive images:



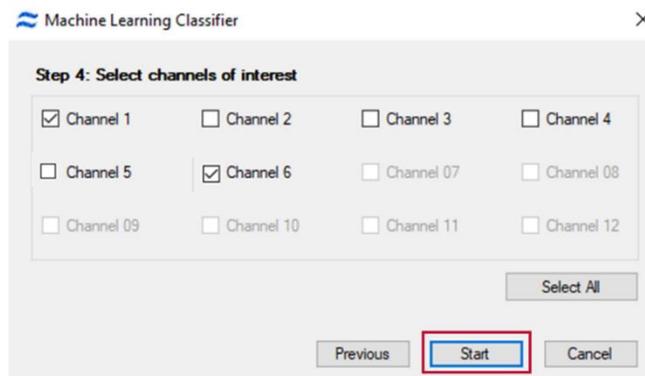


Step 3 allows for the selection of categories that will best differentiate between the populations of interest. All feature categories can be selected, or a subset can be selected to create more specific classifier features. In addition, the checkbox beside 'Only use existing features' can be selected if the user does not want the ML module to compute any additional masks and features, choosing to only incorporate pre-existing masks and features into the ML classifier feature calculation.

The user may also exclude specific features. To do this, click the checkbox beside 'Select base features to exclude' then click the 'Choose' button to bring up a new window in which specific features can be selected for exclusion by the ML module.



Step 4 allows the user to select all channels, or specific channels of interest. Click the 'Start' button to begin classifier feature calculations.





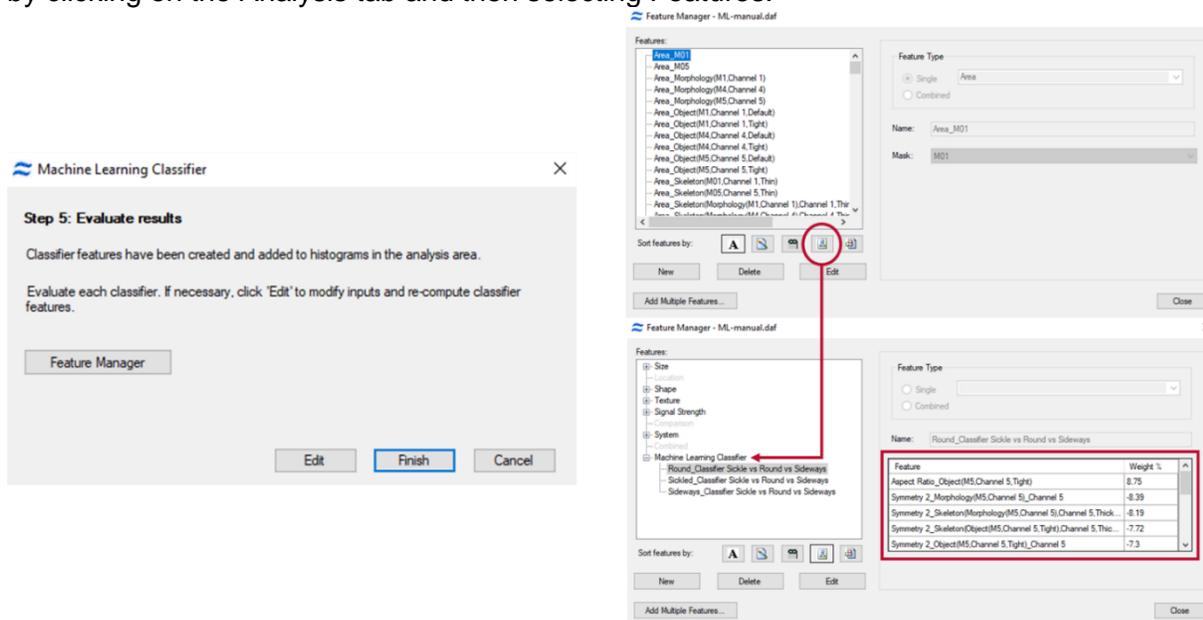
Once the calculations are complete, Step 5 presents the user with options to either edit the selections made in Step 2 through Step 4, to finish the wizard, or to enter the Feature Manager window.

Clicking 'Edit' returns the user to Step 2 and modifications can be made in Steps 2 through 4.

Clicking 'Finish' exits the wizard, prompting the user to save their work.

Clicking 'Feature Manager' brings up a new window with all features displayed. Features are displayed alphabetically by default. Clicking the Feature Category button helps to locate classifier features that have been created by the ML module. Each classifier feature can be highlighted individually which will display a list of all base features used in its creation as well as the weight of each feature.

**Note:** All features in IDEAS®, including classifier features, can always be viewed in the feature manager by clicking on the Analysis tab and then selecting Features.



### 3. Working with classifier features in IDEAS®

Once the classifier features have been calculated, they can be plotted to separate populations of interest. To demonstrate this, the following example examines how to separate round cells, sickle cells and cells that have sideways orientation.

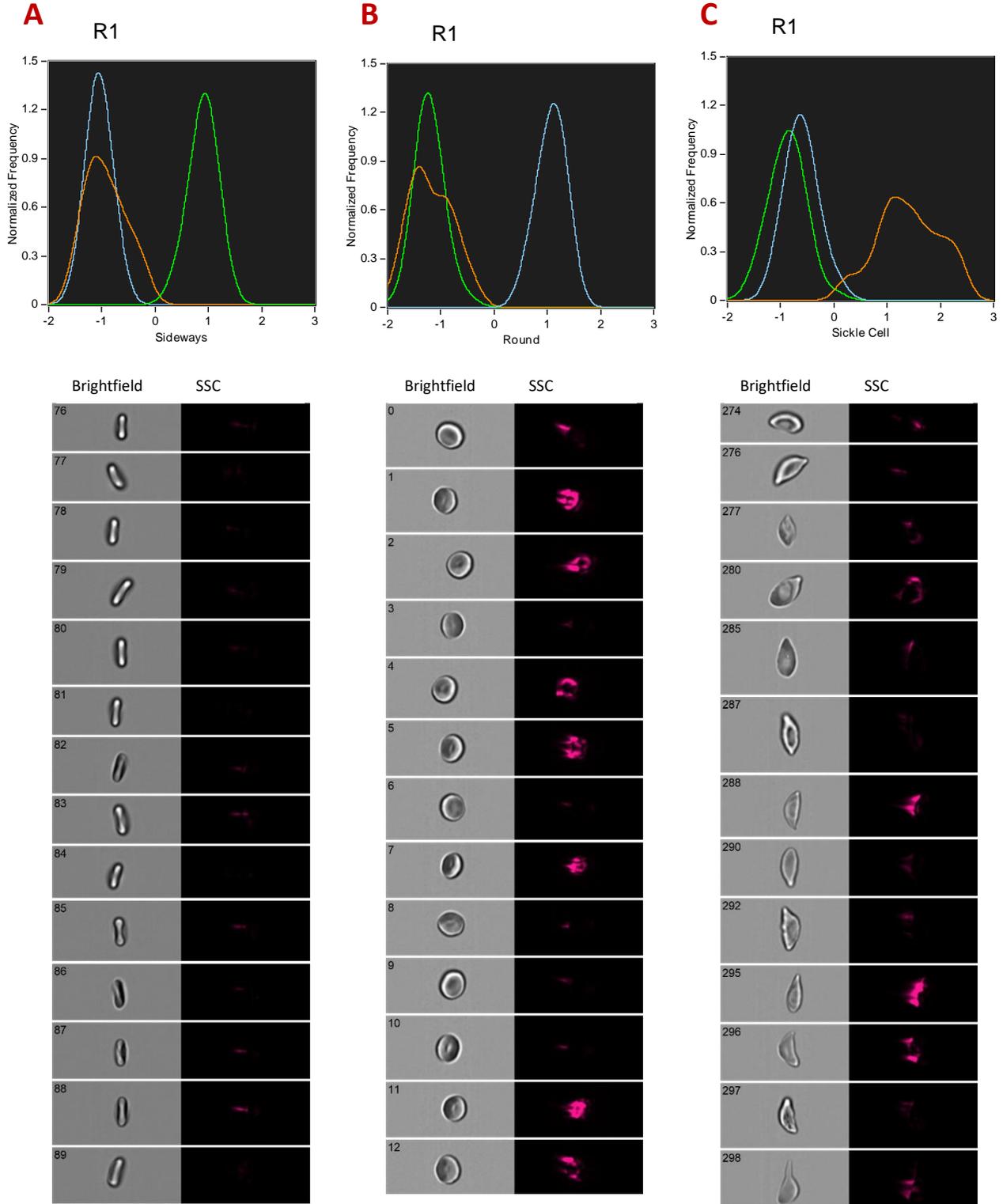
Upon completion of the ML wizard, individual histograms will automatically be generated for each classifier feature that has been calculated using the truth populations.

In this example, three histograms have been created based on the input truth populations; round cells, sickle cells and cells with sideways orientation.

On each histogram, events in bins greater than zero are images that are best represented by the classifier feature.

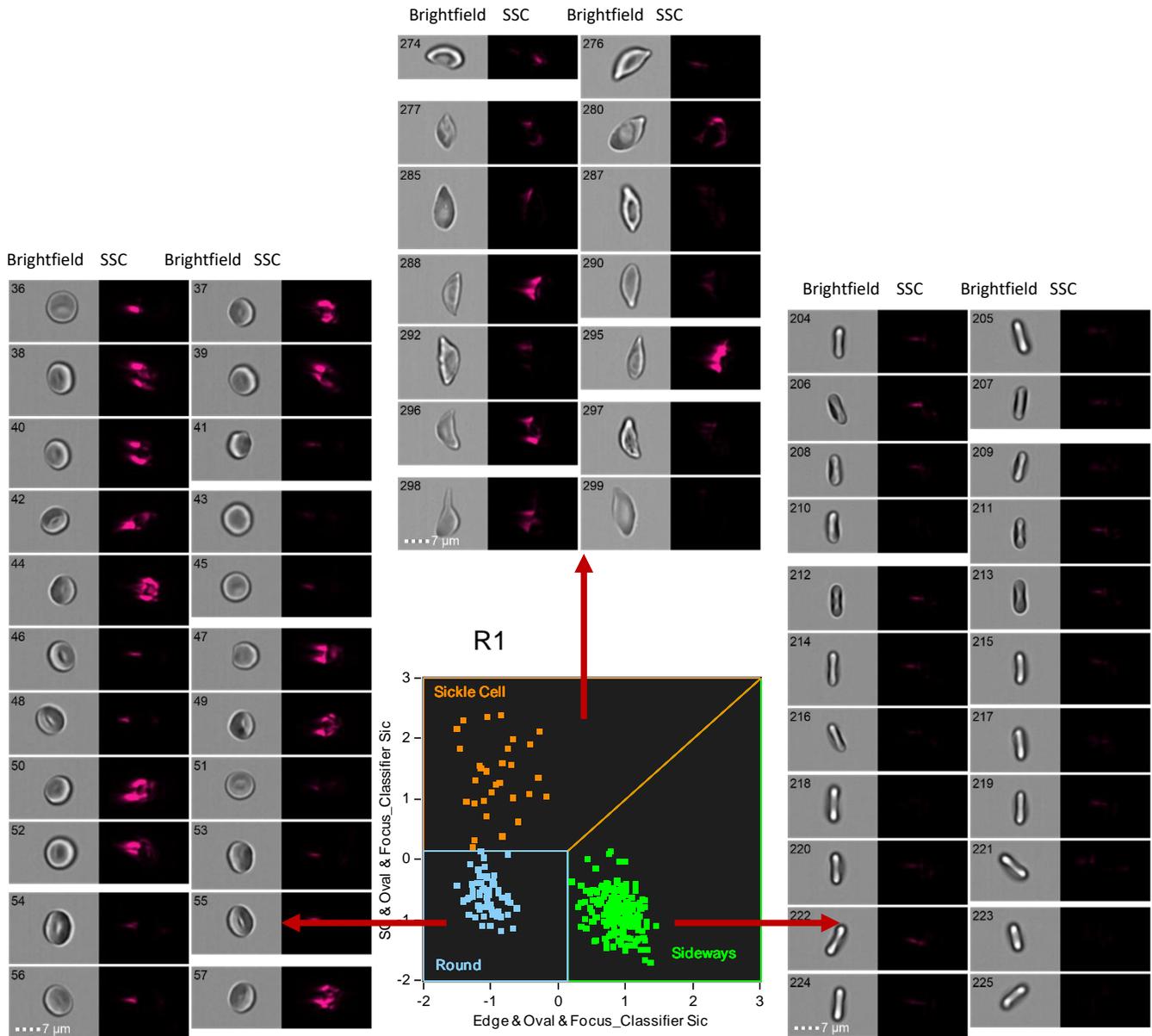


Panel A demonstrates that the cells identified as sideways by the 'Sideways' classifier feature fall between 0 and 3. Similarly, panel B identifies the 'Round' classifier feature, and panel C illustrates the 'Sickle' classifier. Representative imagery from each truth population are shown.





Bivariate plots can also be created using the classifier features to separate populations. In this example, it can be seen that by plotting the Sideways feature and the Sickle feature together, it is possible to separate all three populations of interest on a single plot.



**Best Practices:** Verify results by checking that the classifier feature appropriately identifies the expected phenotype. It may be necessary to tag additional images or use a refined parent population to improve the results. If the classifiers will be applied to other experiment files, it's critical to validate that the classifier feature is using appropriate masking, and features relevant to the desired morphology.