Prediktera.

Breeze Tutorial

Classification of nuts step 1 basic

In this tutorial you will analyse hyperspectral images with samples of nuts (almond, hazelnut, pecan and walnut) and shell for each nut type. The tutorial images contain samples of known nut type that will be used as a training data set, and a sample mix that will be used as a test dataset.

Your goal is to learn how to **use Breeze to make a classification model and then use it to predict the class of new samples**.

Steps included in tutorial		
Record	**Model**	**Play**
1. Download tutorial image data	3. Create a sample model to remove background pixels	5. Create prediction workflows to classify new samples
2. Enter the known class information for the training samples	4. Create a PLS-DA classification model	6. Real time predictions
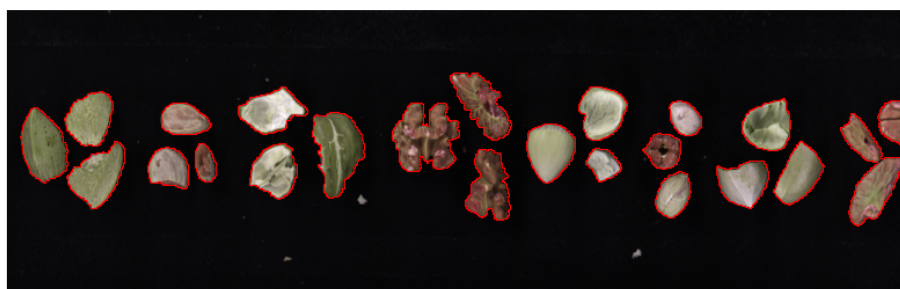
Hyperspectral image
SWIR camera
1000-2500 nm
(data was reduced to
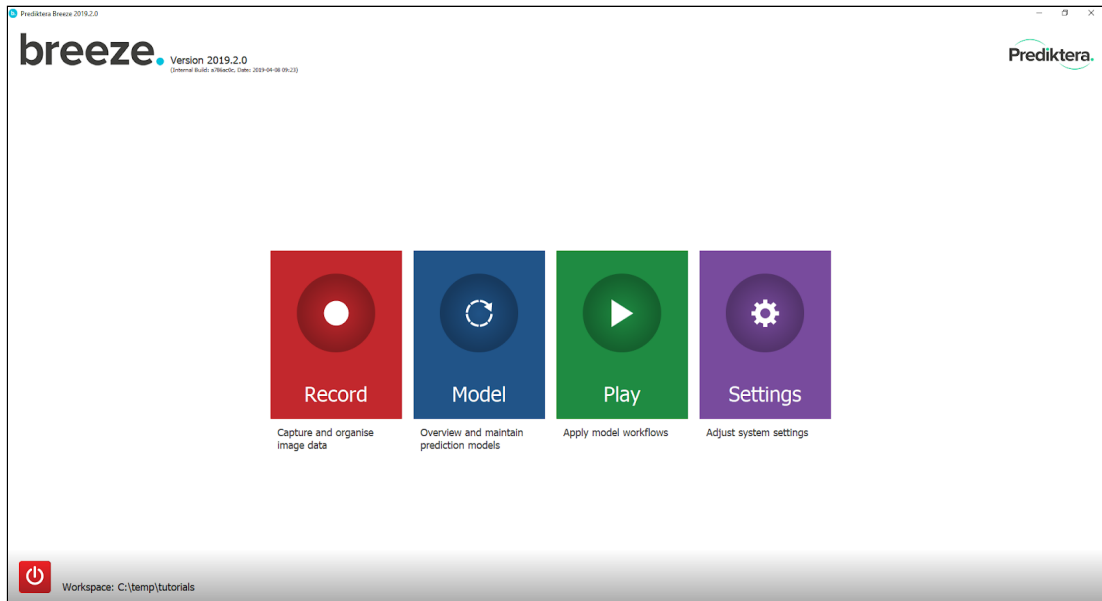31 spectral band to
reduce file size)
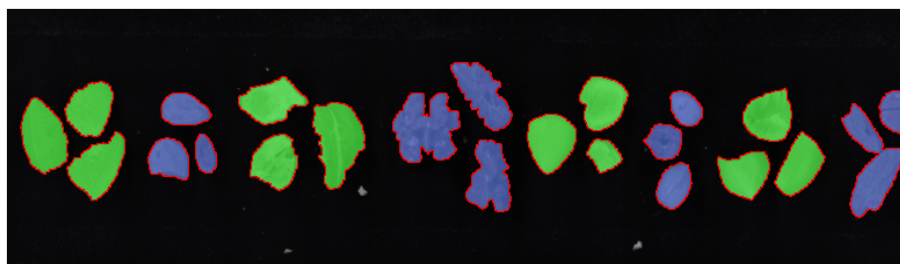
Classification
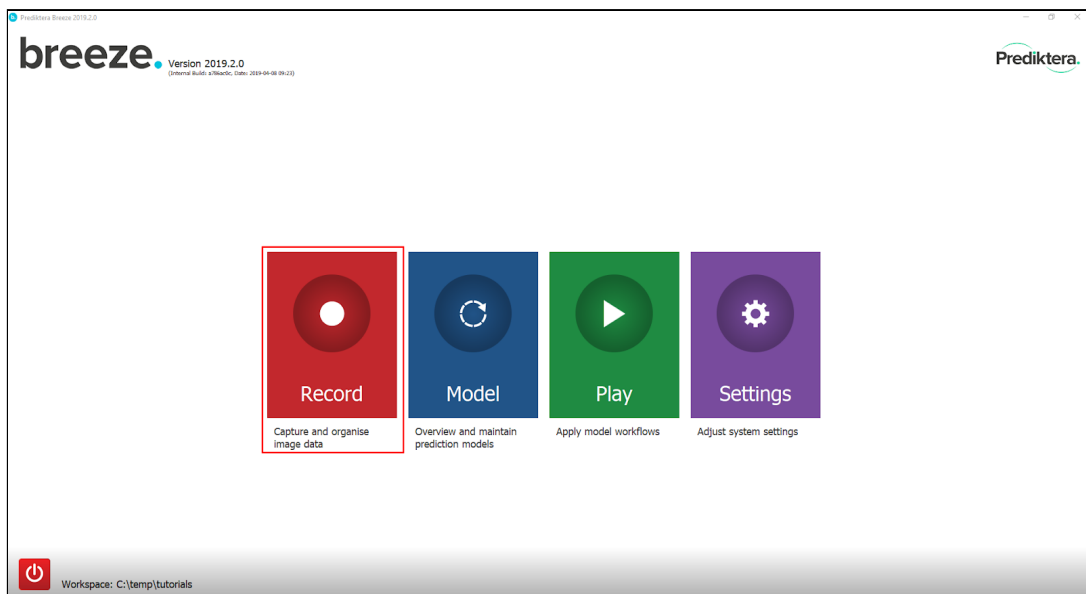🔵 Nut
🟢 Shell

Download tutorial image data

1. **Start Breeze with the shortcut created after installation.**
   The Breeze start screen will look like this:

Breeze is organised into different views depending on the task at hand. Each view has a specific purpose as described under each button.

2. Enter the Record view by **pressing the "Record" button**

3. You will then see the following view (if you already have a Study in Record press the "Add" button in the lower left corner). **Select the "Tutorial"** tab.
Select "Nuts Classification" in the **"Name"** drop down menu.
**Press OK** to start downloading the image data.
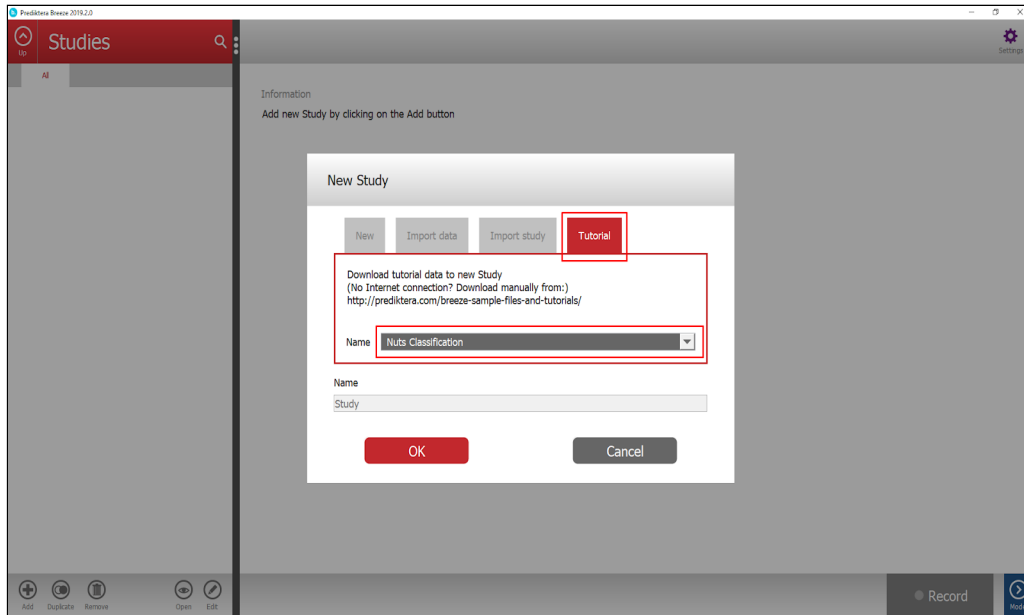
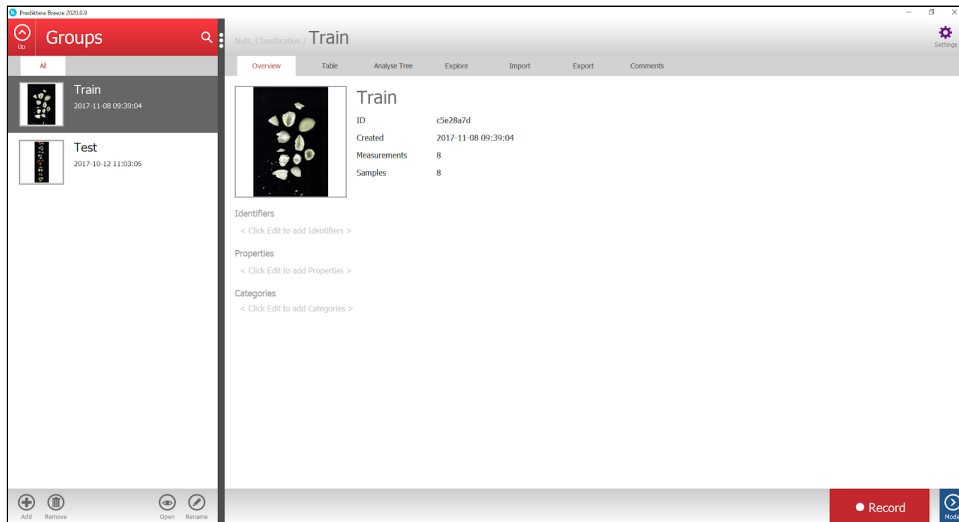4. After the Tutorial data is downloaded you will see the following table:
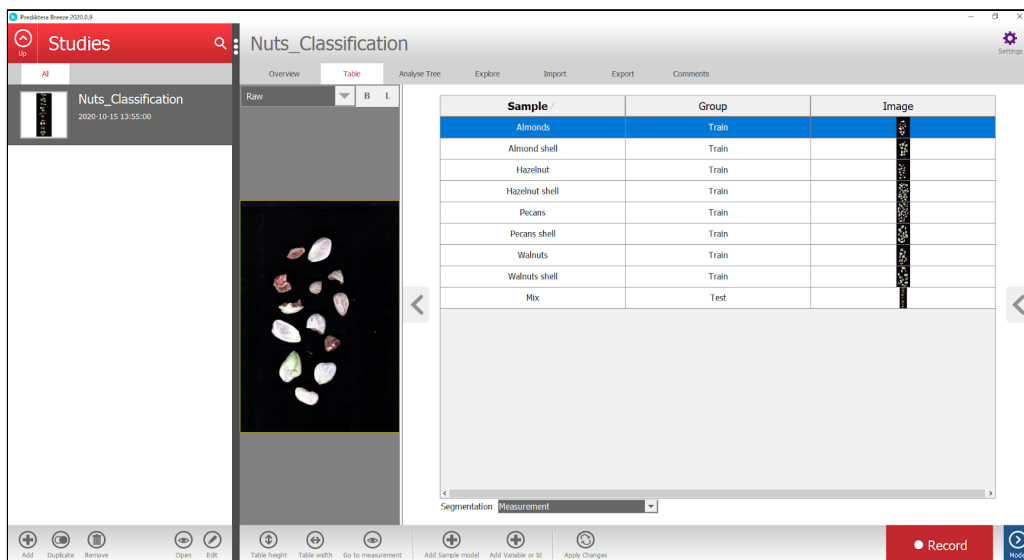
A "**Study**" called **"Nuts_Classification"** has now been created that includes eight training images (with either nuts or shells) and one test image. You can click on a table row to see the preview image (pseudo RGB) for each image.
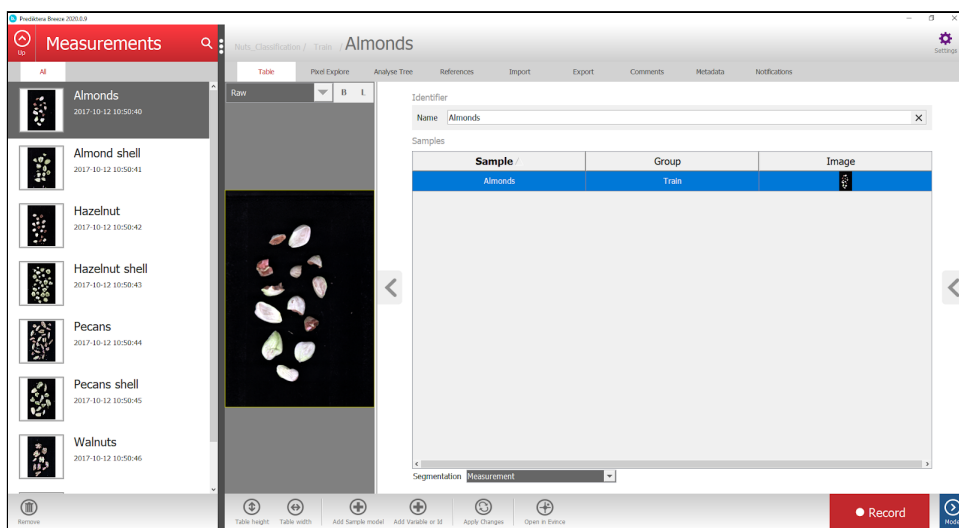
5. **Click on "Open"** to open the study
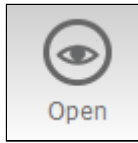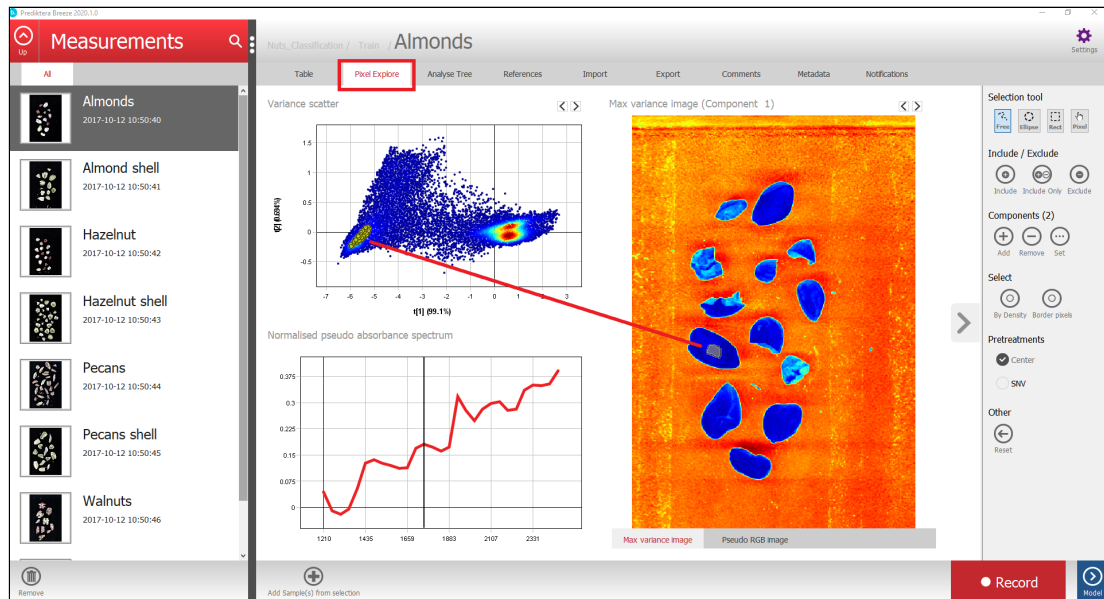

6. The Group view should look like this:

The image data in this study is organized into two **Groups** called "Train" and "Test"

7. **Press the "Open"** button again to open the "Train" group

In the menu on the left side you can now see all the individual images (called **"Measurements"** in Breeze) in this group.

8. **Click on the "Pixel Explore" tab.**

To do a quick analysis of the spectral variation in the image, a PCA model has been created based on all pixels in the image. Each point in the "Variance scatter" plot corresponds to a pixel in the image. The points in the scatter plot are clustered based on spectral similarity. The color of the points in the scatter plot are based on density (i.e. red = many points close to each other).

The "Max variance image" is colored by the variation in the 1st component of the PCA model (the X-axis in the scatter plot, t1), and visualizes the biggest spectral variation in the image. In this case this is the difference between the sample (blue) and the background (red).

Hold down the left mouse button to do a selection of a cluster of points to see where these pixels are located in the image. Move the mouse around in the image to see the spectral profile for individual pixels or do a selection to see the average spectra for several pixels.

9. **Press the "Up"** button in the upper left corner to return to the Group level.
Press "Up" again to go to the Study level.

Enter the known class information for the training samples

1. You will now add the class data to our training data set. Select the "**Table**" tab and **press "Add variable or Id"** (at the bottom of the screen)
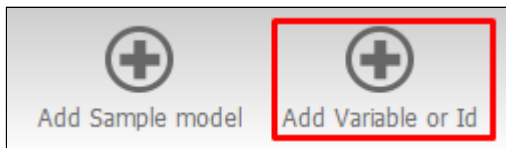
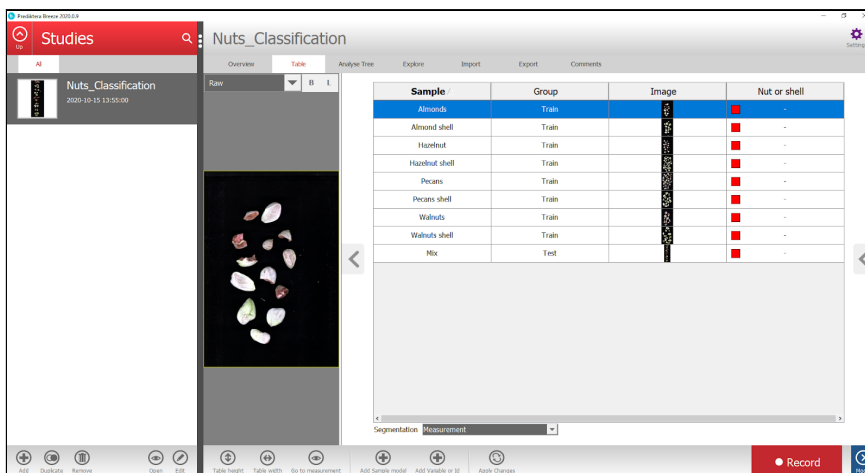2. Select type **"Category (Classification variable)"** and write the name **"Nut or shell"**

3. A column for "Nut or shell" has now been added to the Table

    (If you need to delete variables or IDs, right click on the header for the column you want to delete and then press the "Delete" option that will appear.)

4. **Hold down Ctrl on your keyboard and use your left mouse button to select** the "Almonds", "Hazelnut", "Pecans" and "Walnuts" samples.

   In the "Nut or shell" column **right click on one selected row and write the class name "Nut" and press Enter on your keyboard.**

Sample	Group	Image	Nut or shell
Almonds	Train		
Almond shell	Train		
Hazelnut	Train		
Hazelnut shell	Train		
Pecans	Train		
Pecans shell	Train		
Walnuts	Train		
Walnuts shell	Train		
Mix	Test		

5. **Select the samples with shell and give these the class name "Shell"**

Sample	Group	Image	Nut or shell
Almonds	Train		Nut
Almond shell	Train		-
Hazelnut	Train		Nut
Hazelnut shell	Train		-
Pecans	Train		Nut
Pecans shell	Train		-
Walnuts	Train		Nut
Walnuts shell	Train		-
Mix	Test		-

Your table should look like this now. You do not need to set a class to the Test image.

Sample	Group	Image	Nut or shell
Almonds	Train		Nut
Almond shell	Train		Shell
Hazelnut	Train		Nut
Hazelnut shell	Train		Shell
Pecans	Train		Nut
Pecans shell	Train		Shell
Walnuts	Train		Nut
Walnuts shell	Train		Shell
Mix	Test		-

Create a sample model to remove background pixels

1. You will now create a sample model that will be used to remove the background pixels and to automatically identify the objects (nut samples) in the images.
Press "Add Sample model" button at the bottom of the screen:
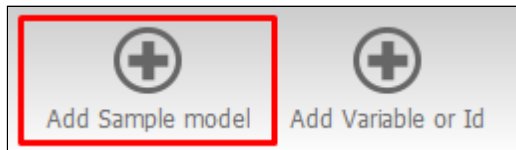
2. **Write a name** for the Sample model (or just use the default name) and **press "OK"**

3. In the first step of the sample model wizard you can select the images that you will use in the model. By default all measurements are included, which is ok.
Press "Next"
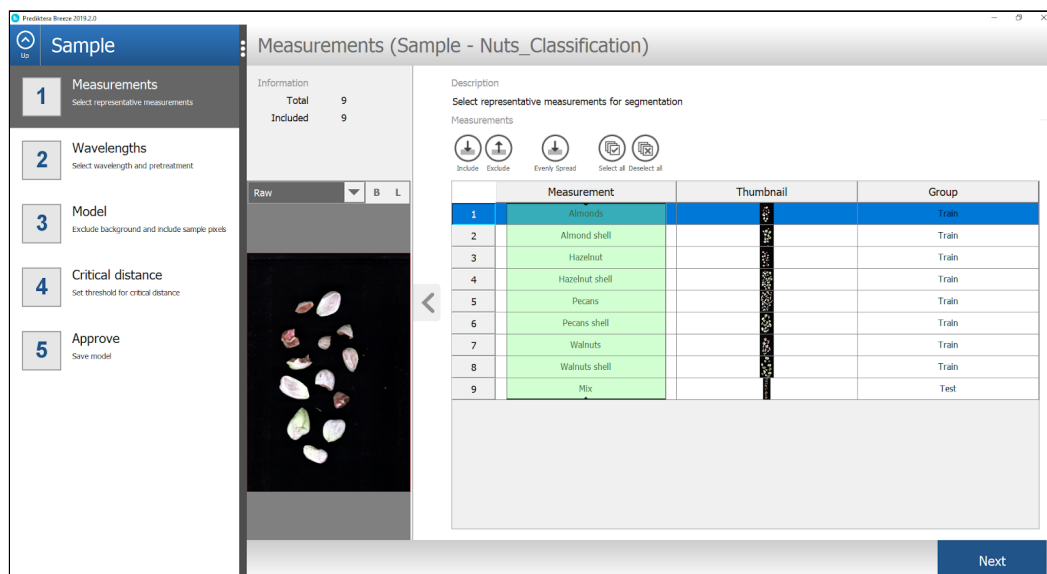
4. In the next step of the wizard you can select spectral bands (wavelengths) to use in the model. By default all wavelengths are included except the first and last bands (the bands at the start and end are often more noisy and 4 % of bands are therefore excluded by default in Breeze).
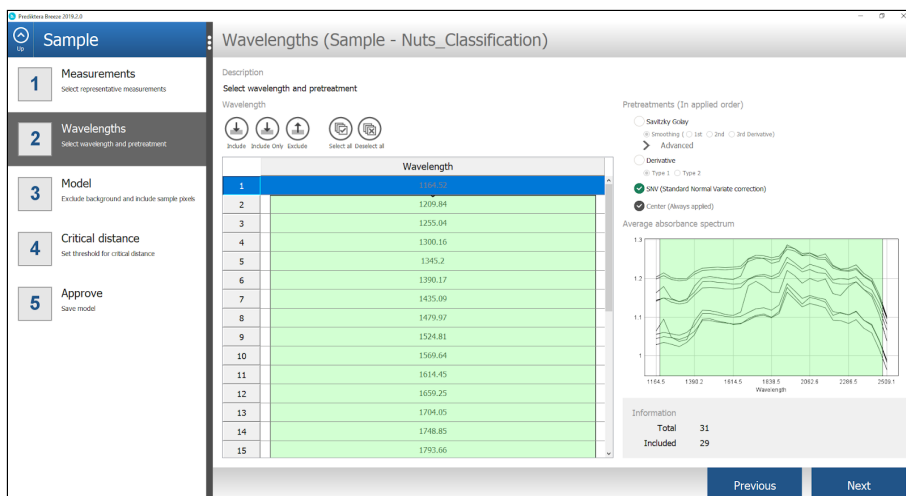
The default is OK so **Press "Next"**.

5. In the next step you will select the pixels to use in the sample model.

   a. A mosaic has been created of all images and a PCA model has been created from all the pixels in this mosaic. **Select a region containing only nut or shell pixels** by holding down the left mouse button and marking the area inside one of the objects. To make this easier you can use the mouse scroll wheel to zoom in.
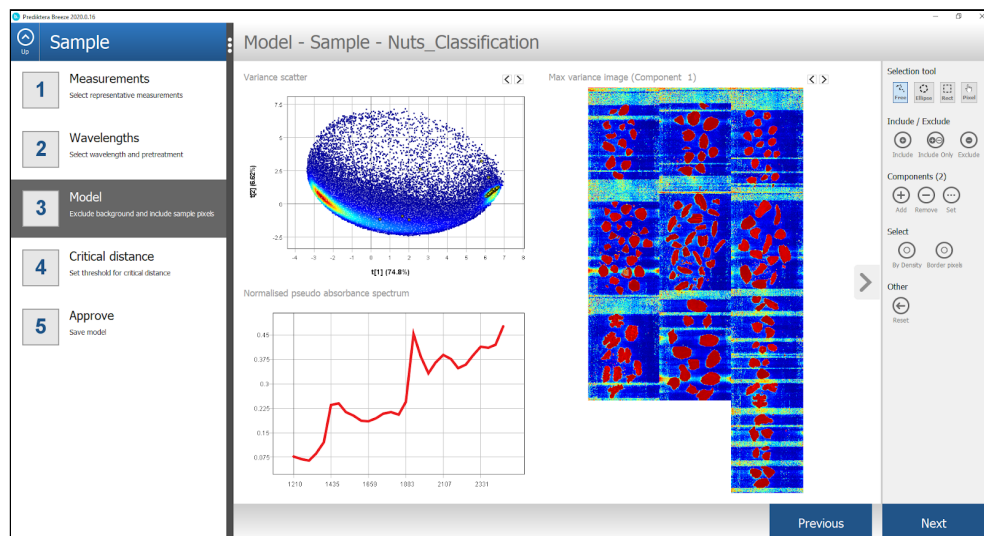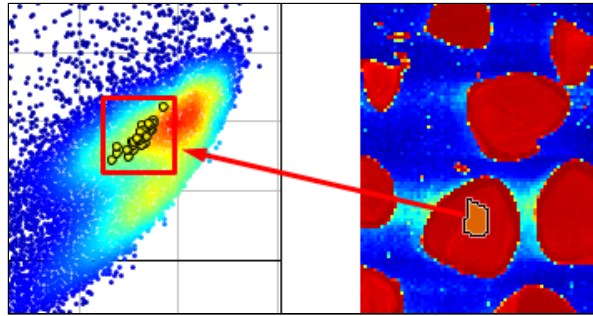
The corresponding pixels are then selected in the scatter plot to the left (see selected pixels inside the circle). Now you know that the nut and shell pixels are in the cluster on the right side in the Scatter plot.

b. **In the scatter plot, select all pixels in the cluster on the right side** (use the pixel density coloring red, yellow, green and light blue as help)**. The corresponding pixels are then selected in the image on the right side and should correspond to all nut and shell pixels.
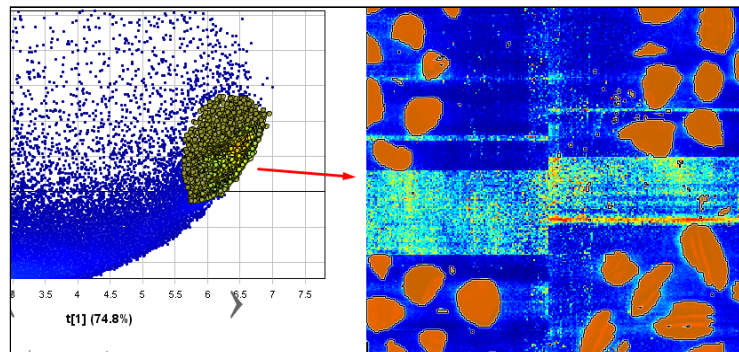
c. **Press the "Include Only"** button in the menu:

d. The plots are now updated and will contain mostly the nuts pixels.
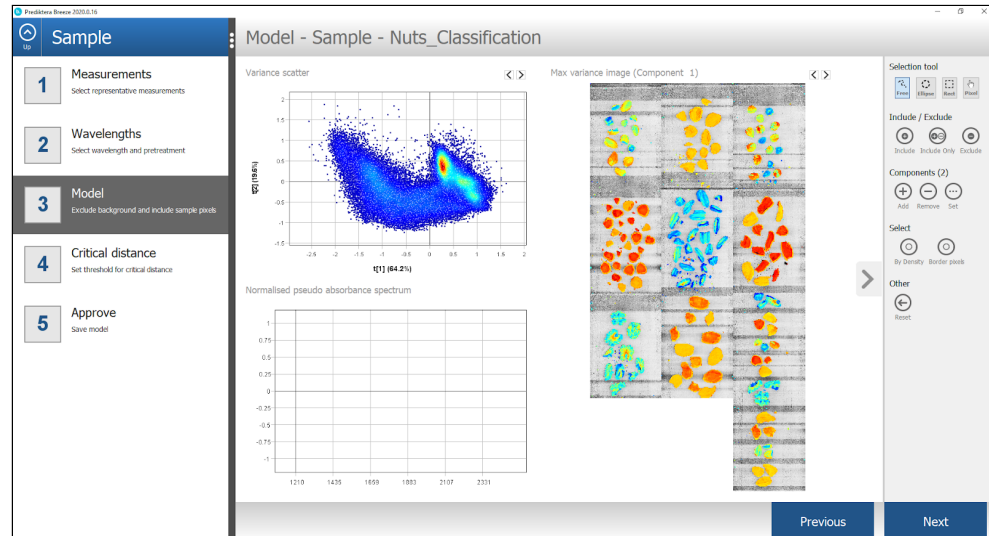
e. To clean up the nuts pixels even more you can remove the pixels bordering each sample object.
Press the "Select - Border pixels" button.

f. Use the default of 1 border pixel and **press "OK"**. The border pixels have now been selected.
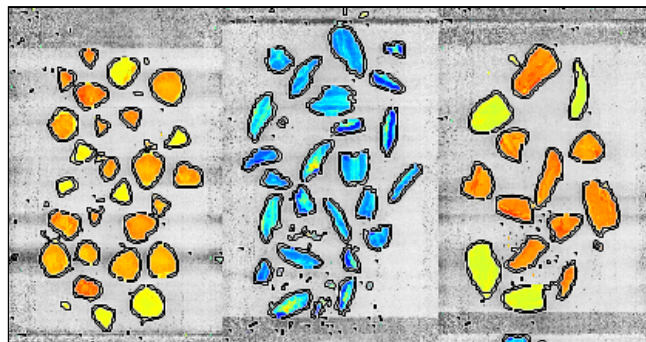
g. Exclude the border pixels by **pressing the "Exclude" button**
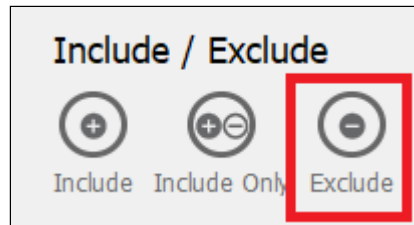
h. **Press Next**

6. In the next step you will set the Critical Distance threshold. This is the distance to the sample model, and will be used to determine if pixels are sample or not. The histogram is showing the distance to the model for all pixels in the images. Pixels on the left side of the red vertical line (critical distance) are inside the threshold.

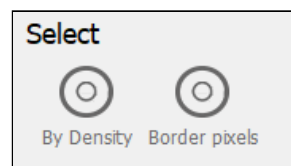   **Drag the red line to the right to move the threshold**. As you can see from the image, more pixels are included when doing this. The aim is to find a level where **all nut pixels are included but not pixels from the background.**
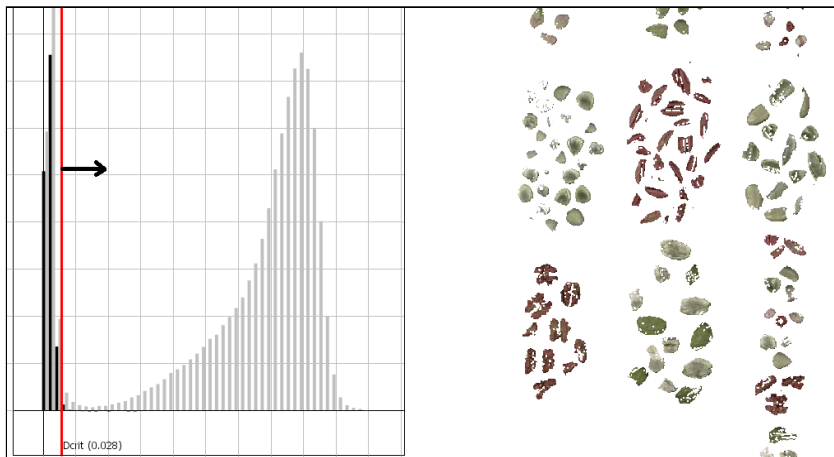
As a general recommendation you can drag the red line to the bottom of the "valley" between the sample and the background bars in the histogram as shown below.

---
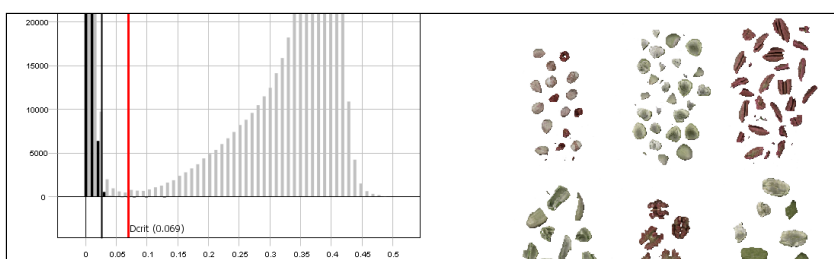
7. **Press "Next"**

8. You are now at the last step of the sample model wizard. The "Minimum area size" is used to automatically exclude smaller unwanted objects (for example dust or dirt). Breeze calculates a suggested minimum area size for your data. In this example any objects under 300 pixels will be excluded from the image. Depending on how you did the pixel selection in the previous step this value might vary. A value around 300 should be OK.

   Press "Finish" to create the sample model and apply this to all images in the study.
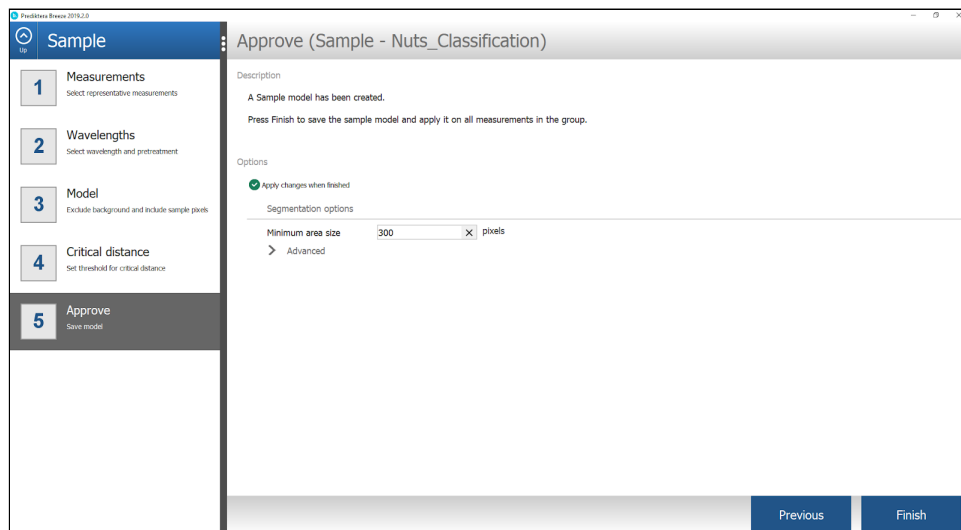
9. In "Table" for the study you can now see all the sample objects in the images after the sample model has been applied and the background pixels removed.
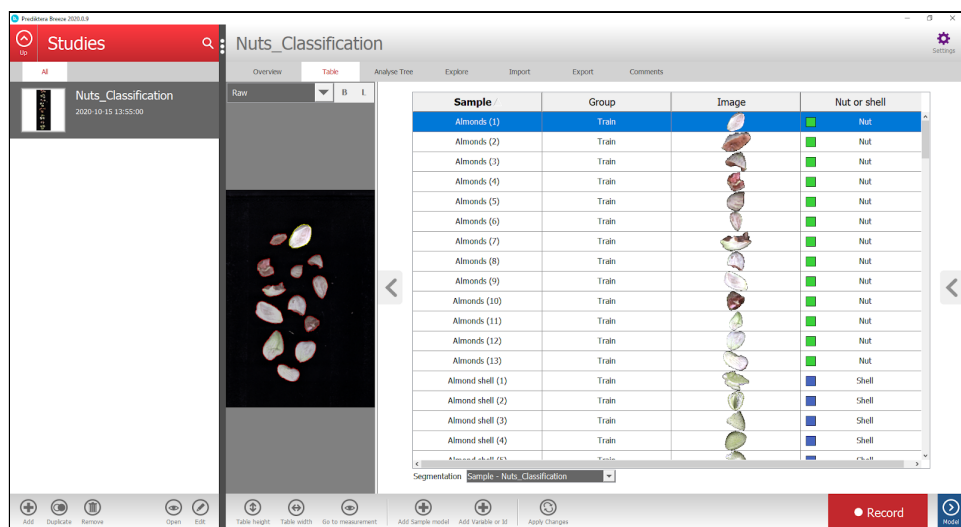
Click on a sample in the "Nut or shell" column in the table to color all objects in the preview image based on the class. You can also click on the objects in the preview image to see where they are in the table.

10. **Press the "Explore" tab.** A PCA model has been created based on the average spectrum for each sample. Each point in the scatter plot corresponds to a sample and the points are clustered based on spectral similarity. Select one or several points to see their average spectrum.

In the menu on the right **click on the category name "Nut or shell"** to color the scatter plot and preview image based on the different classes (the red dots are the "Mix" samples where we had not entered any class).
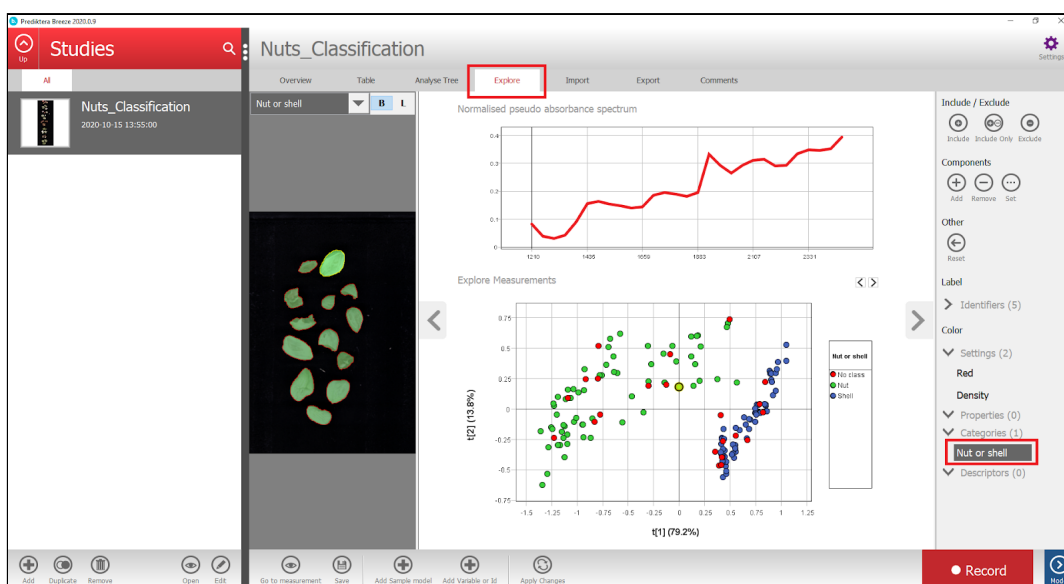
By pressing the arrow you can hide or show the preview image. By pressing the three dots on the vertical band you can expand your screen.
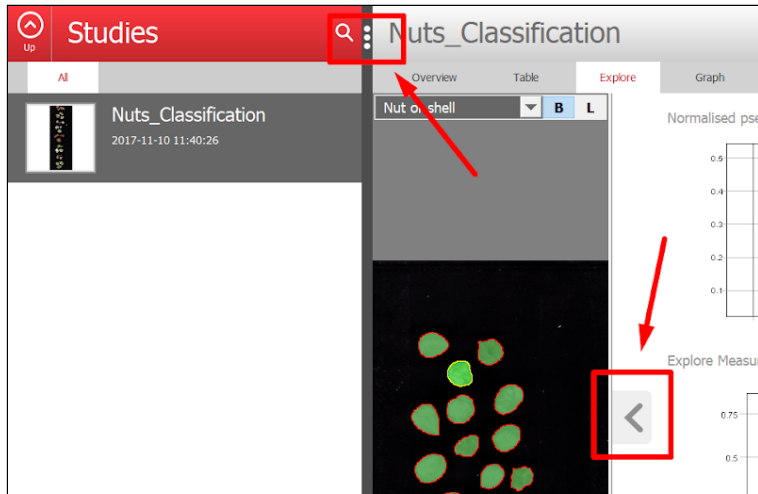
It will then look like this:
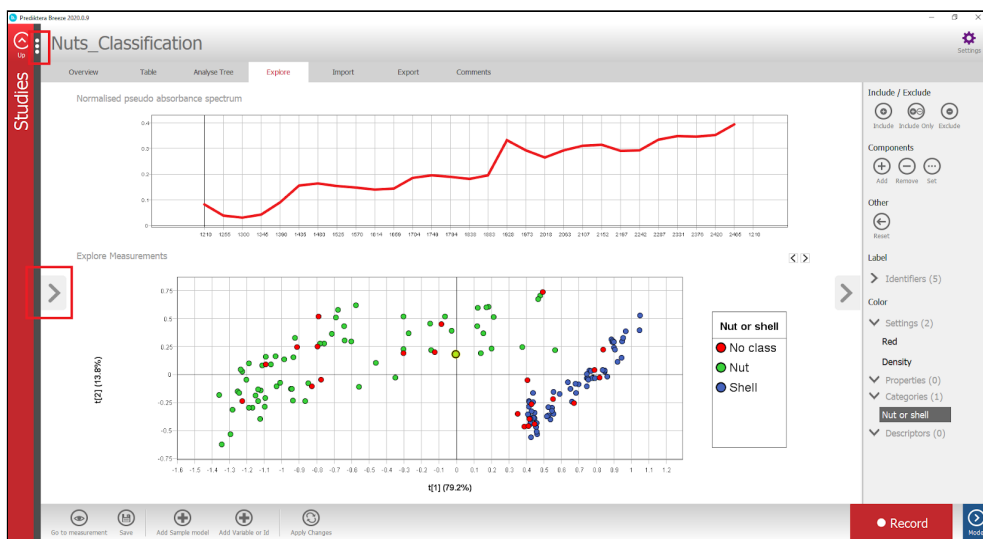
Create a PLS-DA classification model

1. You will now use the average spectrum for each sample and the class type that you have set to train a classification model.
   Press the "Model" button in the lower right corner of the screen to move to the Model step.
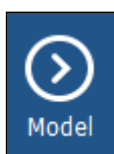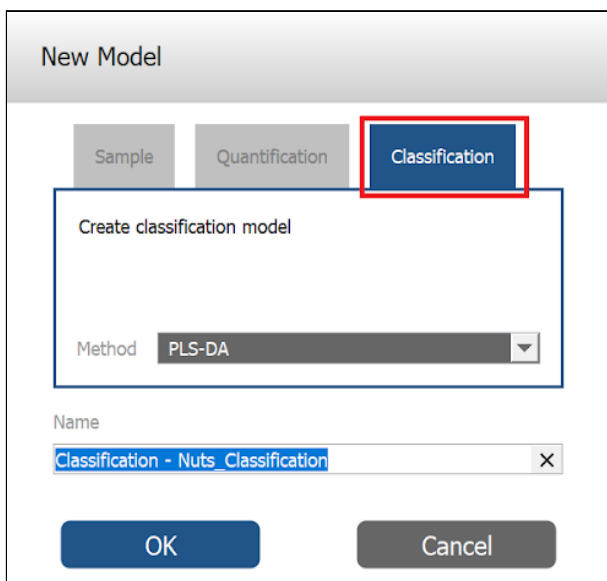
2. In the menu on the left you can see the Sample model that you created before. To make an additional model **Press the "Add"** button.
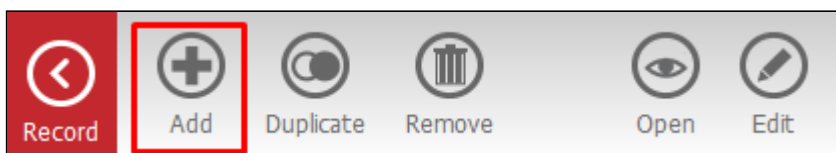
3. In the window that appears, **press the "Classification"** tab, and then **select "PLS-DA". Write a name** for the model (or just use the default name) and **press "OK".**

4. In the first step of the Classification wizard you can select the Category (class Y-variables) that you will use to build the model. Since we only have one in the study you can just **press "Next".**
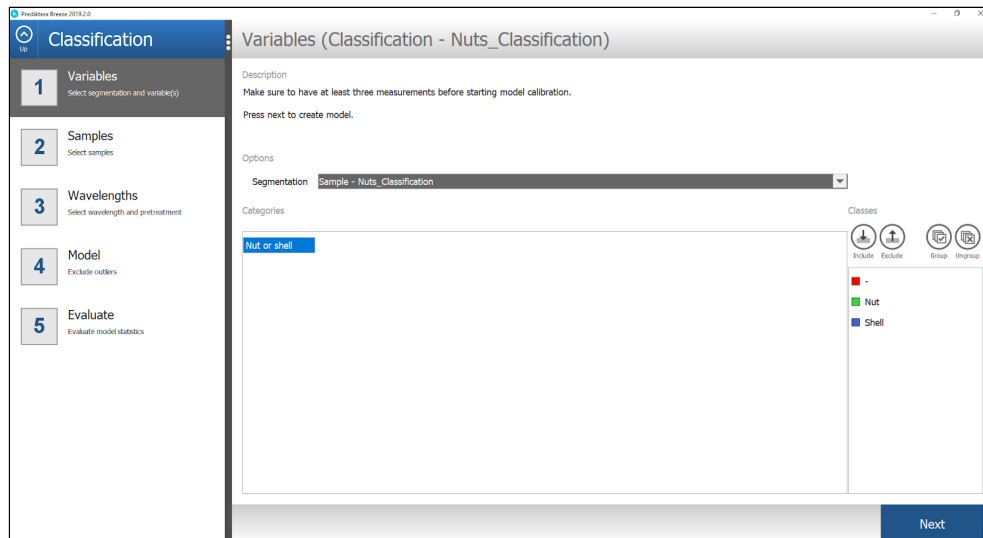
5. In the next step of the wizard you can select the samples that you want to include in the model. By default the measurements from the "Train" group have been included since they have entered class information. The measurement "Mix" in the "Test" group has been excluded since there was no class information entered. This is OK. **Press "Next"**.
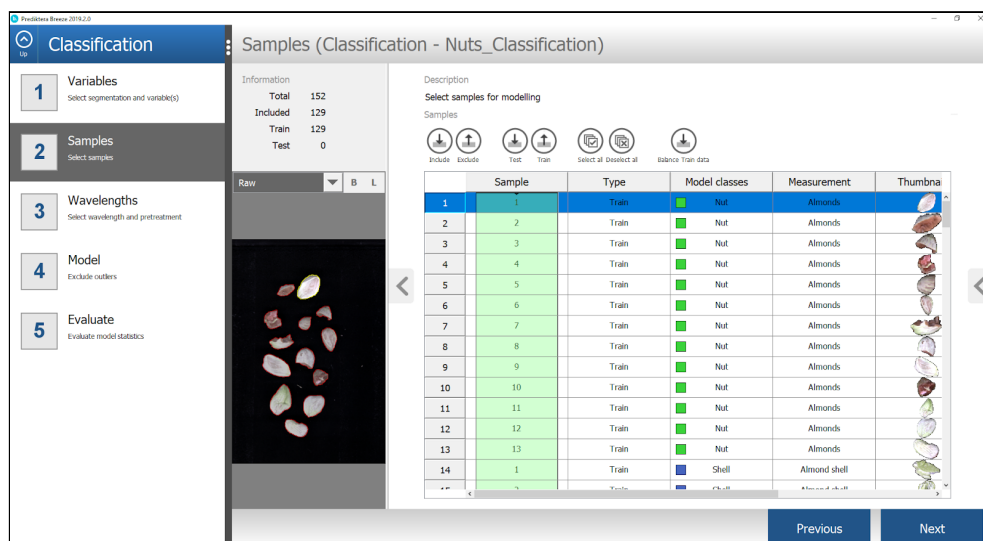
6. By default all wavelengths bands are included except the first and last bands. The graph on the right is showing the average spectrum for each sample. Above this graph there is an option to select different pretreatments of the spectral data. By default "SNV" is used. All default settings here are OK.
Press "Next".

7. A PLS-DA classification model has now been calculated.

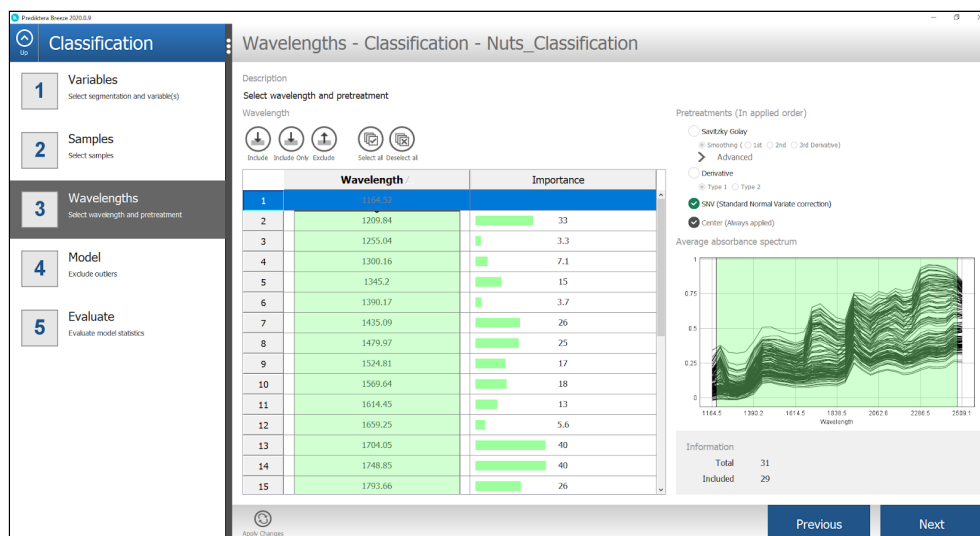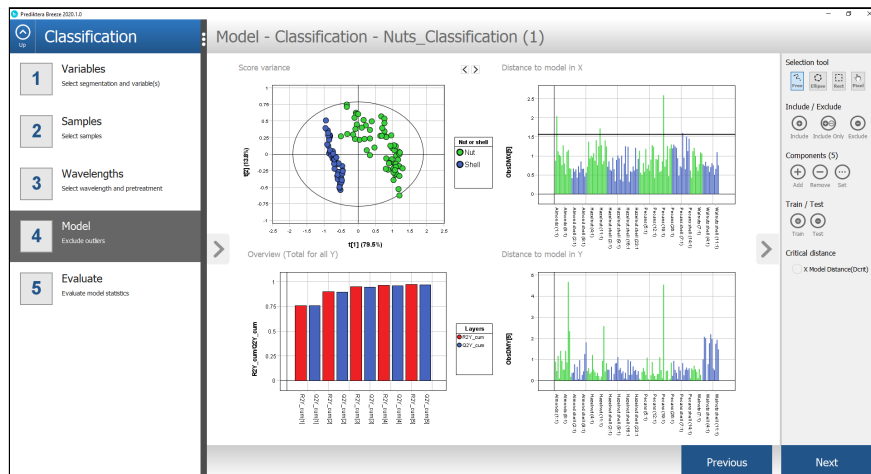a. The "Overview (Total for all Y)" graph is showing how good the PLS-DA model is. It also shows the number of components used for the model. In this case the autofit used five components. The R2 (model fit) and Q2 (prediction from cross validation) using five components are around 0.97 indicating a very good model. An R2 and Q2 value of 1.0 indicates a model explaining all the variation. A value of 0 indicates that no variation can be explained.

b. The "Distance to model in X" and "Y" graphs show the distance to the model for each sample. A high bar indicates that the sample might be an outlier (for the X distance the horizontal black line can be used as a guide. The "Model" scatter plot and the "Distance to model" graphs can be used to identify and exclude outliers.

Since everything looks OK **Press "Next"**.

8. In the last step of the wizard you can evaluate how good the model is. The "Nut or shell" vs. "Ycalc.Nut or shell" is showing how good the model can separate the two classes. The "Variable overview" is showing the R2 and Q2 for each class. Everything looks OK.

Press "Finish" to complete the model.
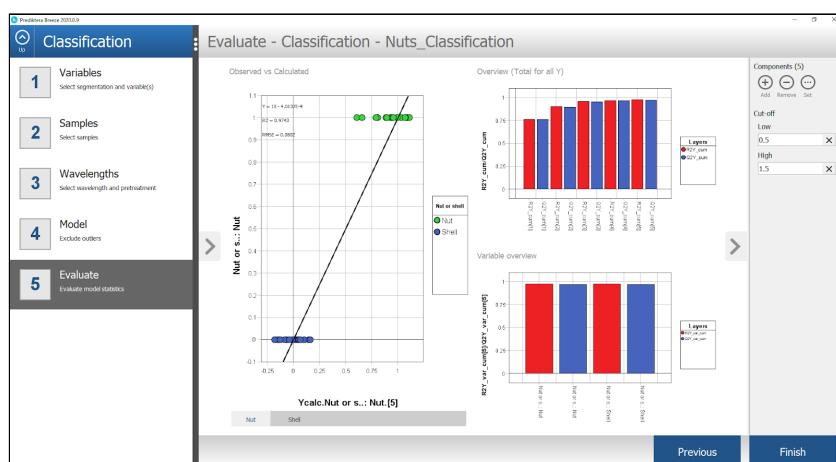
3. The Classification PLS-DA model has now been saved as you can see in the menu on the left. With this model selected **press the "Classification"** tab to see how many samples were correctly classified.
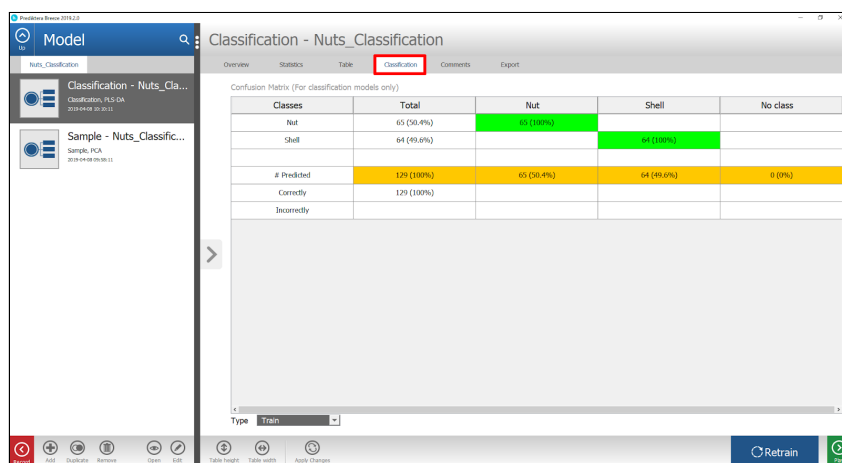
Each row is showing the correct class for the samples, and the columns are showing the classes that these samples are classified into by the model. In this example all the samples are correctly classified.
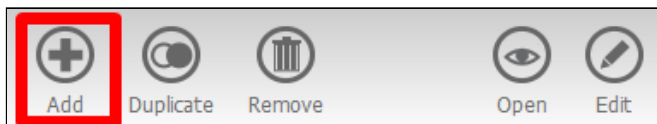
Create prediction workflows to classify new samples

1. In this step you will use the classification model to analyse the image with unknown samples.
Press the "Play" button in the lower right corner to move from the Model mode to the Play mode.
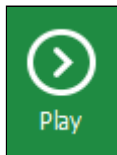
2. **Press the "Add" button** to make a new workflow

3. In the window that appears, **select the "Record data"** tab and **select the "Test" Group**. Write a name for the workflow or just use the default name.
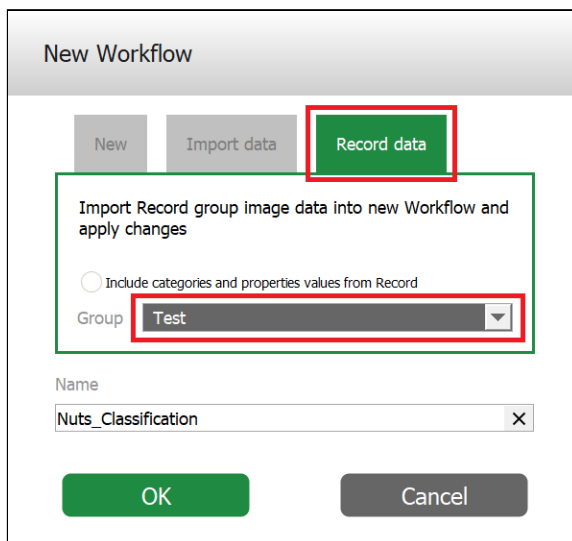
A new workflow will be generated based on the models you have created for this study (sample and classification model). The images from the "Test" group in Record will be imported and applied to this workflow.
**Press OK.**

4. A table is generated with the predicted class of all samples in the "Test" image.
   Click in the "Nut or shell" column to color the preview image based on the class.
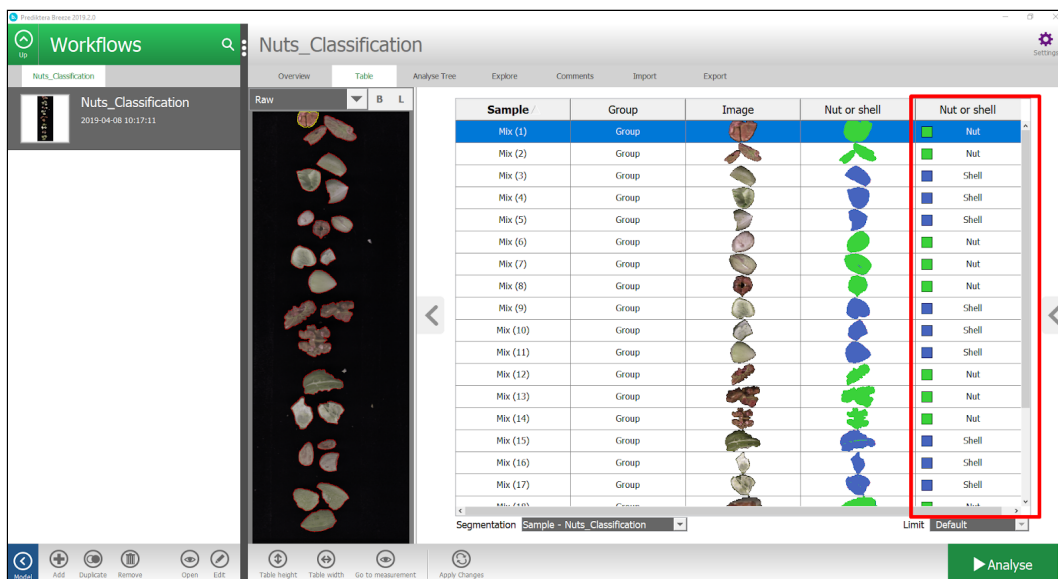

The "Nut or shell" column with the colored square (as shown above) is showing the class for the sample based on its average spectrum. The "Nut or shell" column with the small thumbnail image of the sample is showing the classification based on the spectrum for each pixel.

5. **Click on the "L" button above the preview image** to add a legend with classes.
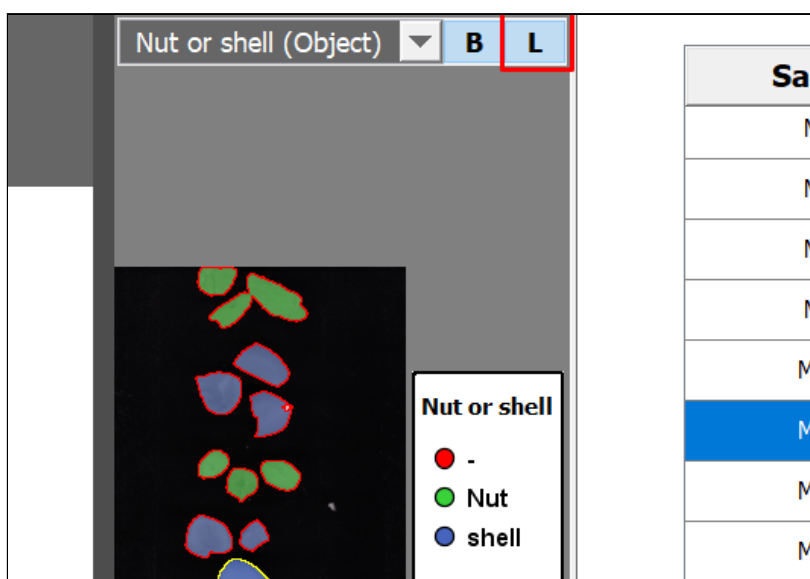
6. **Press the "Analyse Tree" tab to see the steps in the "Workflow".** First, the "Measurement" (image) is analysed by your sample model ("Sample - Nuts_Classification") to find the sample "Object".

For this object it then applies your classification model to calculate the variables ("Nut or shell").
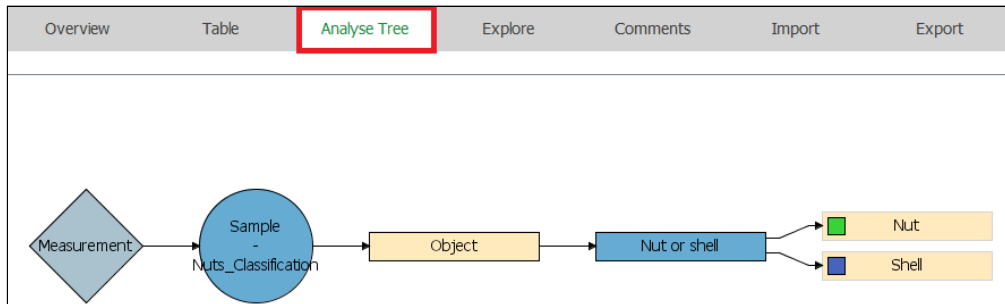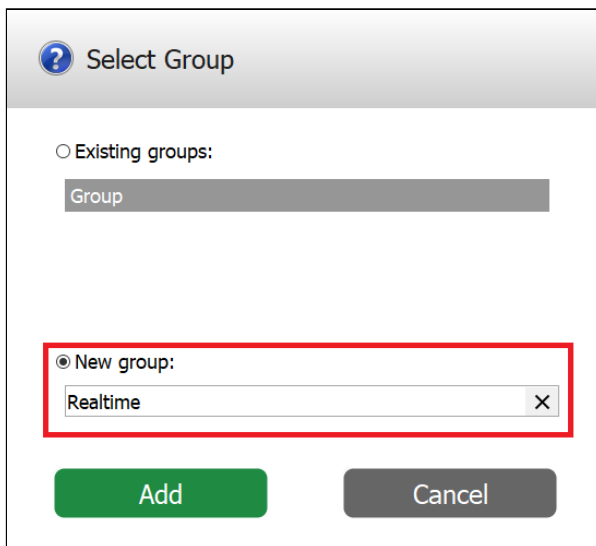
Real time prediction

1. In addition to analysing images that are already recorded on your harddrive, you can also use Breeze to analyse images in real time directly from the camera. If your computer is not connected to a camera, you can simulate this by using the camera simulator in Breeze. With this it will read images from your harddrive and analyse them continuously. By default it will use the measurements from the current Record study as input.

   With your workflow selected, **Press "Analyse"**
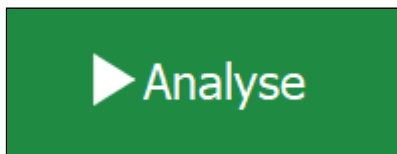
2. In the window that appears **select "New Group", give it a name** like "Realtime" and **press "Add".**
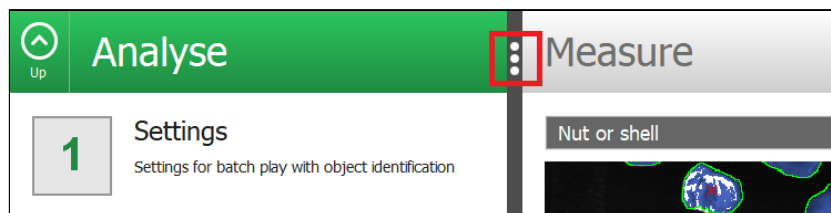
3. In the next step of the Play wizard you can select if you want to save the image measurement and the calculated descriptors (in this case the predicted class), and the raw data (spectral data for all pixels) for the images being analysed. If you are scanning many images you can uncheck this option to save space on your hard drive. You can also set how the real time image segmentation should be done (Parallel and Sequential)
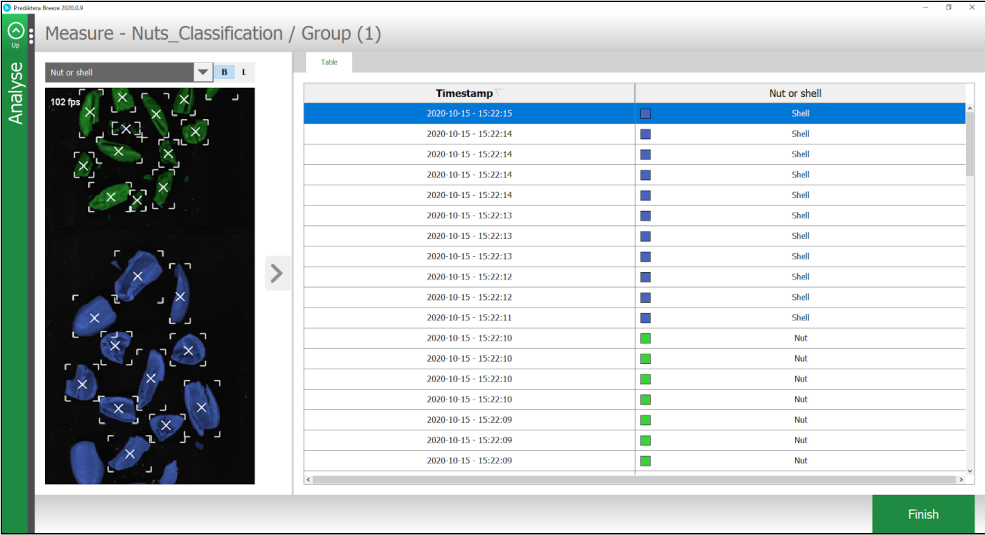
Press "Next".
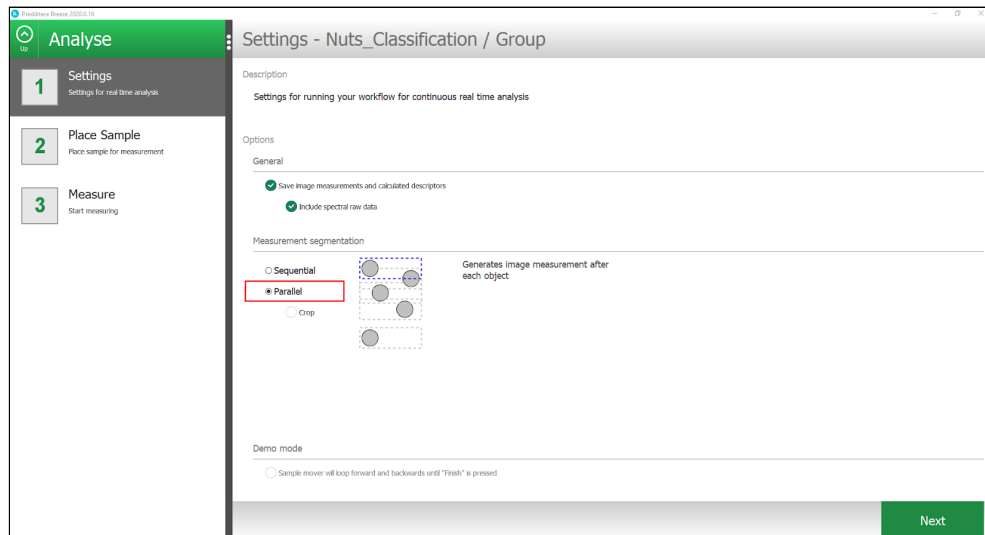
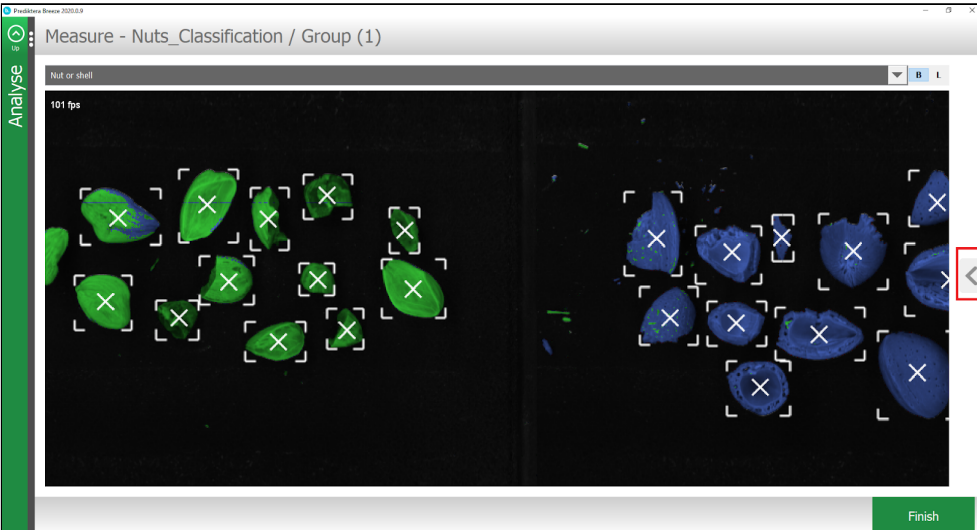4. **Press the three dots on the vertical band** to enlarge the image.

5. As you can see the image is analysed in real time and the results are displayed in the table.

Press the button with the arrow pointing right to make the real time image full screen.

Nice job! You have reached the end of the step 1 of the "Classification of Nuts" tutorial.
If you would like to learn more about classification analysis please try the tutorial step 2 to learn some additional features.

"Classification of nuts – step 2 advanced"