

Breeze Tutorial

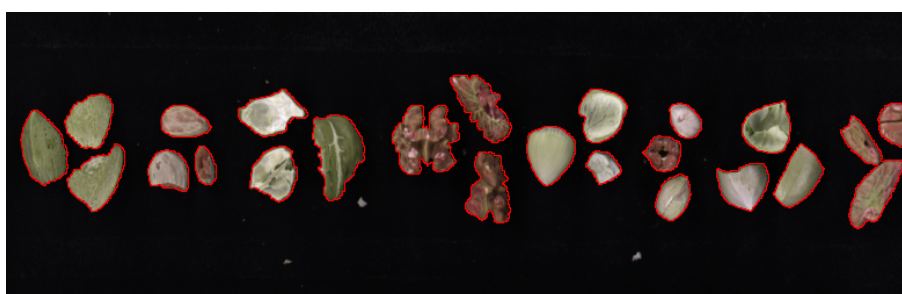
Classification of nuts step 2 advanced

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 contains samples of known type and an unknown mix of samples.

Before doing this tutorial, first do the tutorial “Classification of nuts step 1 - basic”.

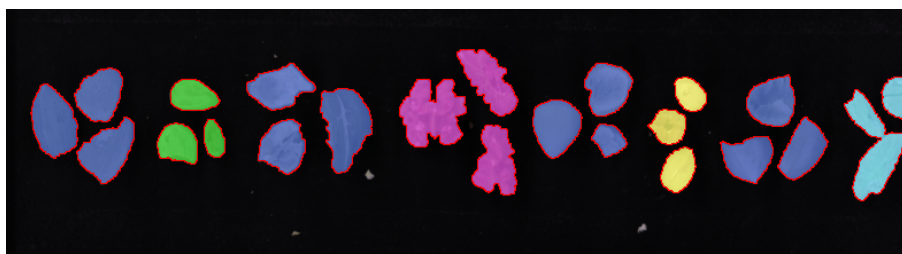
Your goal is to learn how to use **Breeze** to make a classification model and then use it to predict the class of new samples.

Steps included in tutorial		
Record	Model	Play
1. Import the known class information for the training samples	2. Create classification model (PLS-DA)	5. Create workflow and Import Record test data
4. Import the known class information for the test samples	3. Create classification model (SIMCA)	



Hyperspectral image

SWIR camera
1000-2500 nm
(data was reduced to 31 spectral band to reduce file size)



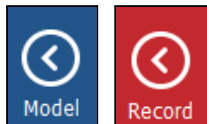
Nut type

- Almond
- Shell
- Hazelnut
- Pecan
- Walnut

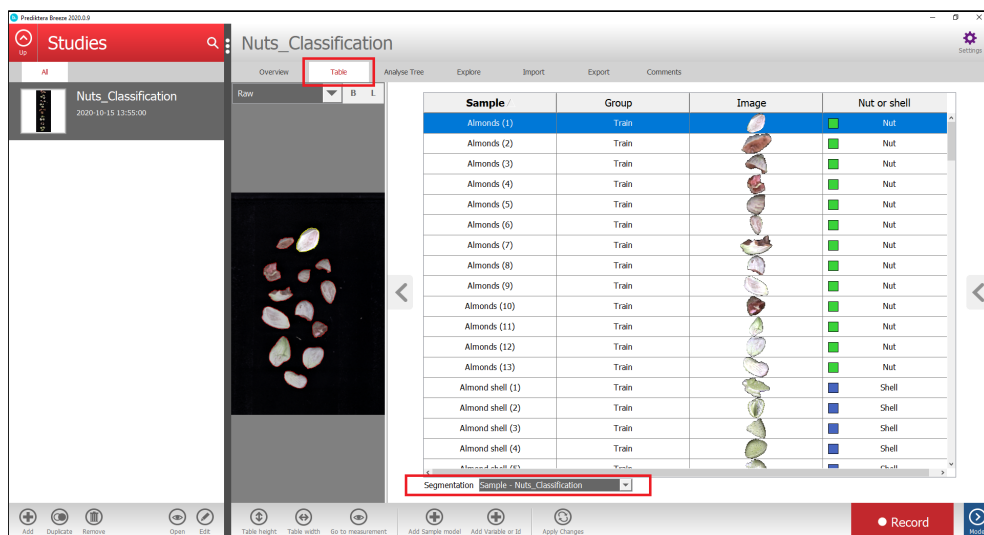
Import known class information to samples

In this tutorial you will use the same images as in the tutorial “Classification of nuts step 1 - basic” but add a new class variable to classify the type of nut. You will also test two classification models types (PLS-DA, SIMCA).

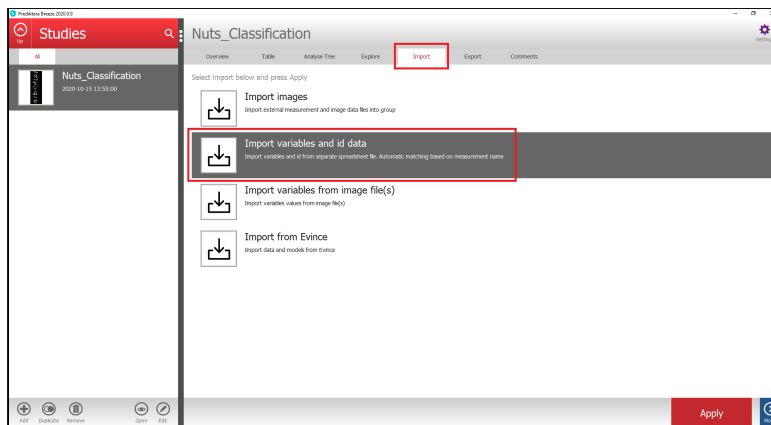
1. Make sure that you are in the Record mode (if you are in the “Play” or “Model” mode use the shortcut buttons (lower left corner) to return to the Record mode).



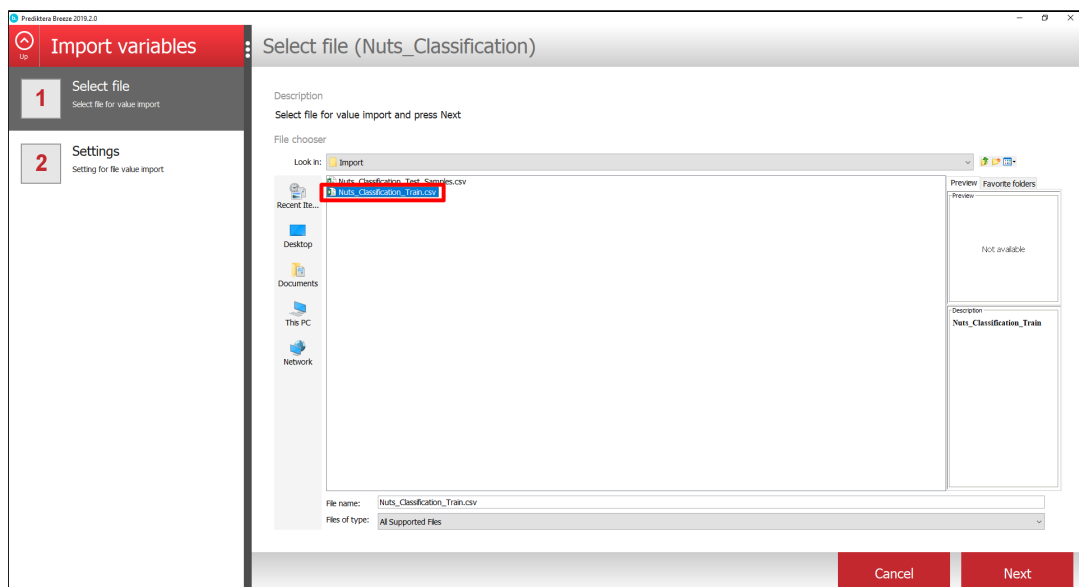
The “**Nuts_classification**” study and “**Table**” tab should be selected. Make sure that the **Segmentation** drop down menu under the table is set to “**Sample - Nuts_Classification**”.



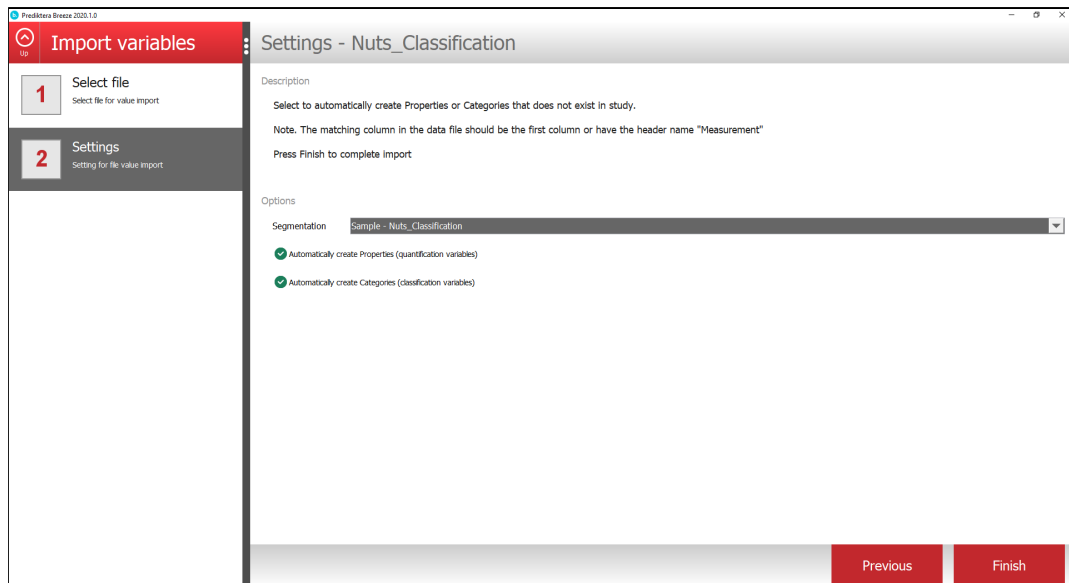
- Press the “Import” tab, select “Import variables and id data” and press “Apply”.



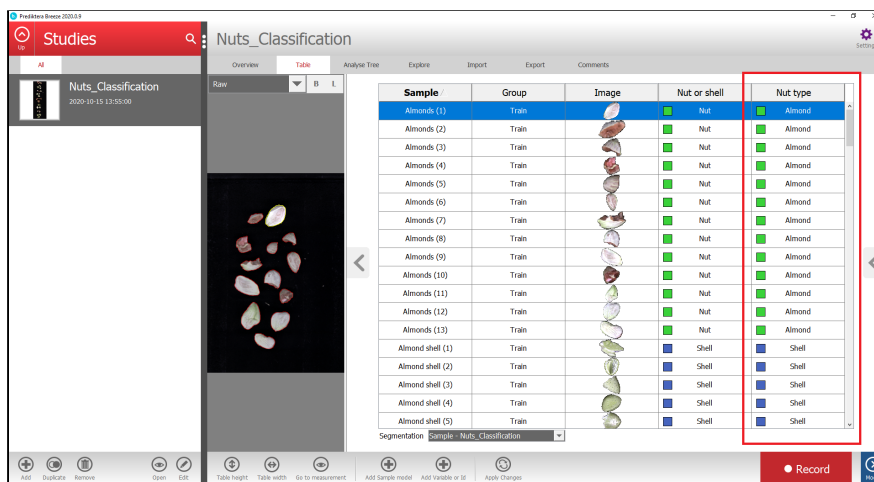
- Select “Nuts_Classification_Train.csv” and press “Next”.



- Make sure the correct segmentation and the option to “Automatically create Categories” are selected and press “Finish”.



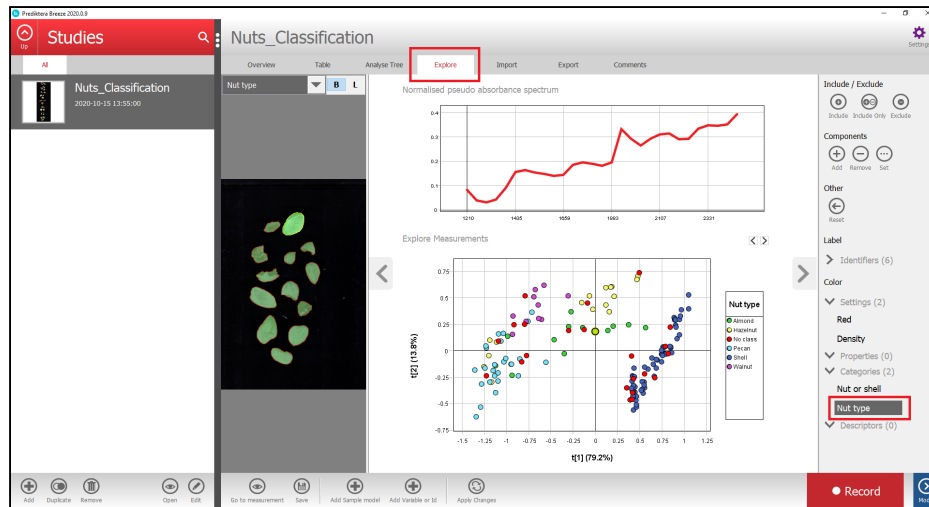
5. In the Table view you should see the new “Nut type” class variable that was imported. The reference values were automatically matched with the correct sample object.



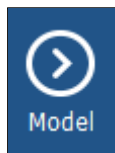
The spreadsheet .CSV file that you imported looks like this when opened in Excel. The column “Measurement” matches the class data (“Nut or shell” and “Nut type”) to the correct images and samples.

	A	B	C	D
1	Measurement	Nut or shell	Nut type	
2	Almonds	Nut	Almond	
3	Almond shell	Shell	Shell	
4	Hazelnut	Nut	Hazelnut	
5	Hazelnut shell	Shell	Shell	
6	Pecans	Nut	Pecan	
7	Pecans shell	Shell	Shell	
8	Walnuts	Nut	Walnut	
9	Walnuts shell	Shell	Shell	

- Press the “Explore” tab and then under “Color” select “Nut type” to see how the different types cluster

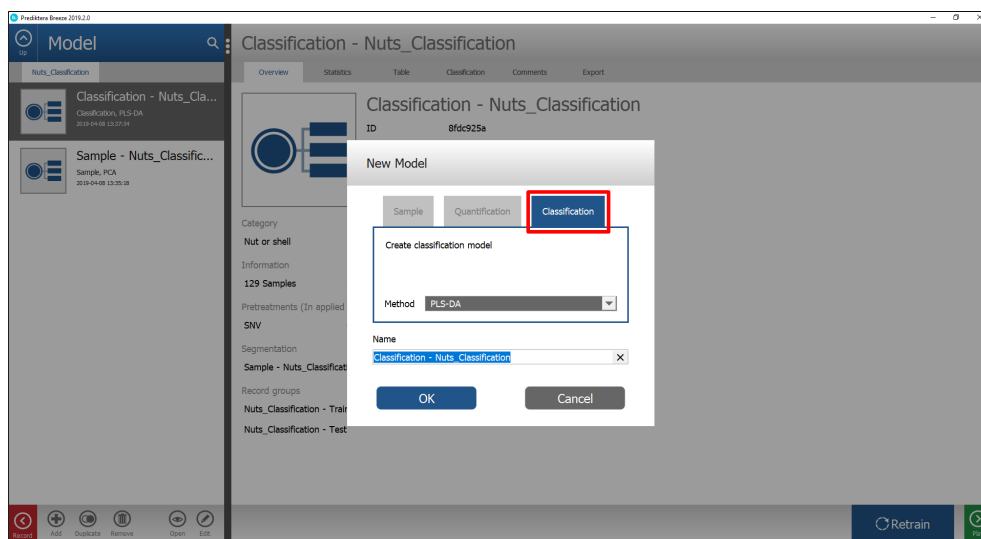


- Press “Model”

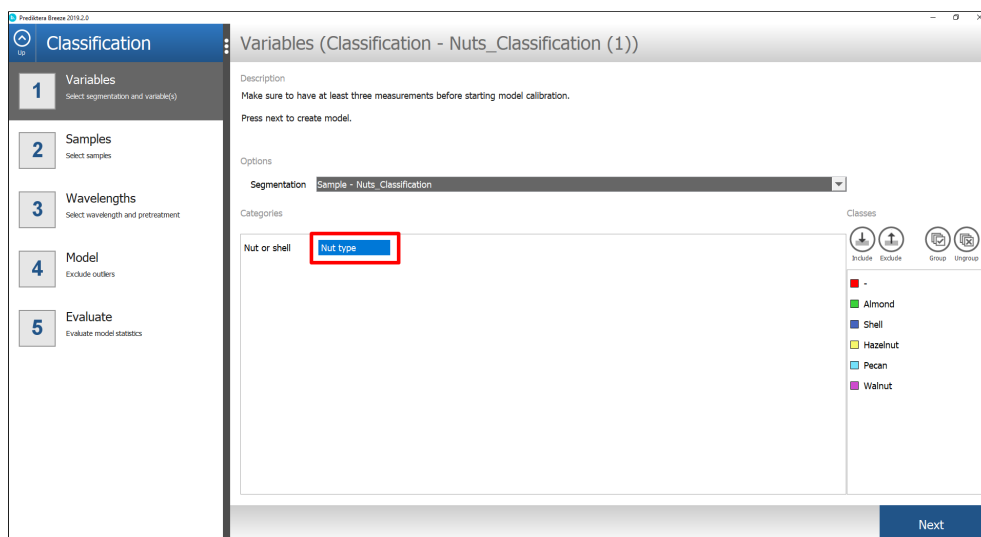


Create classification model (PLS-DA)

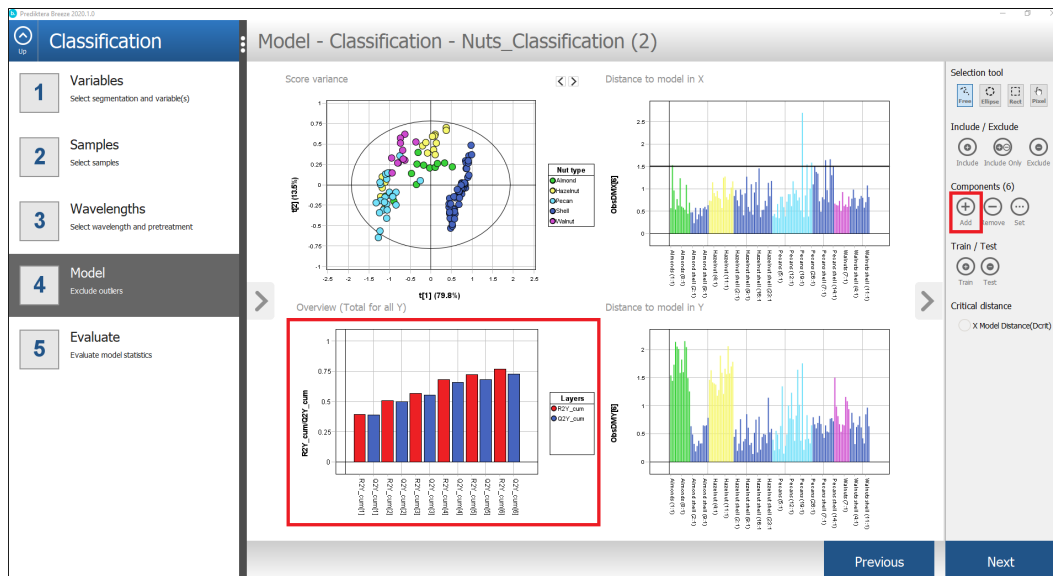
- Now you will create a Classification model for “Nut type”. Press **Add** to make a new model. Select “Classification” and “Method” PLS-DA. Write a name or use the default. Press **OK**.



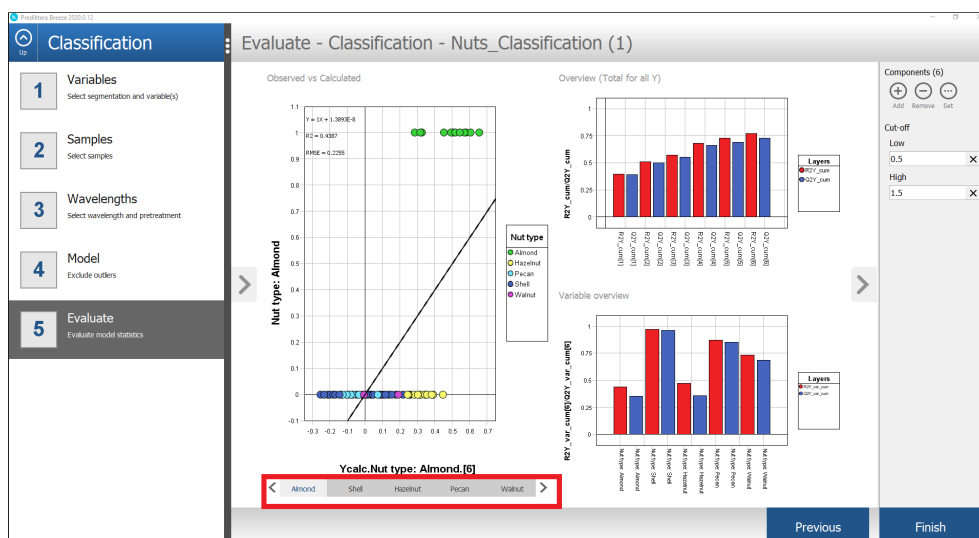
2. Choose the “Nut type” category in the first step of the wizard and press “Next”



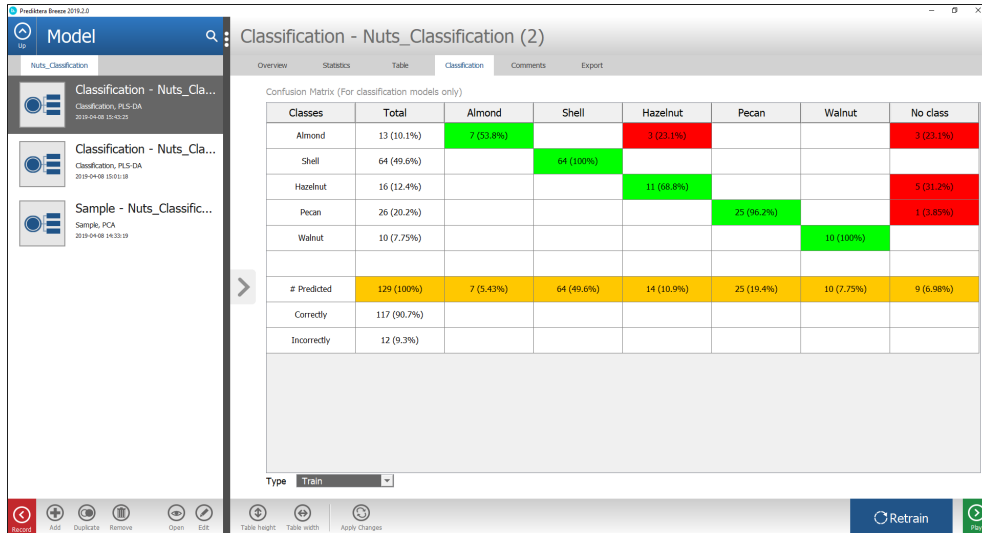
3. In the 2nd and 3rd step of the model wizard just press “Next” (use the default) so that you come to the 4th step (“Model”).
4. In the “Overview (total for all Y) you should have 6 components for your PLS-DA model. Under “Components” you can press “Add” to add more components to the PLS-DA model (to a total of 6, i.e. 6 red and 6 blue bars). In the “Overview (total for all Y)” you can see that the model improved (higher R2 and Q2). Press “Next”.



- In the last step you can see the class separation by pressing the tabs for the different classes under the “Nut type” vs “Ycalc.Nut type” plot. The “Variable overview” graph shows that Almond and Hazelnut could not be classified as well as the other types. Press “Finish”.



6. Press the “**Classification**” tab to see how well the training data was classified by the PLS-DA model (the results might vary slightly depending on how you did your sample model).



Confusion Matrix (For classification models only)

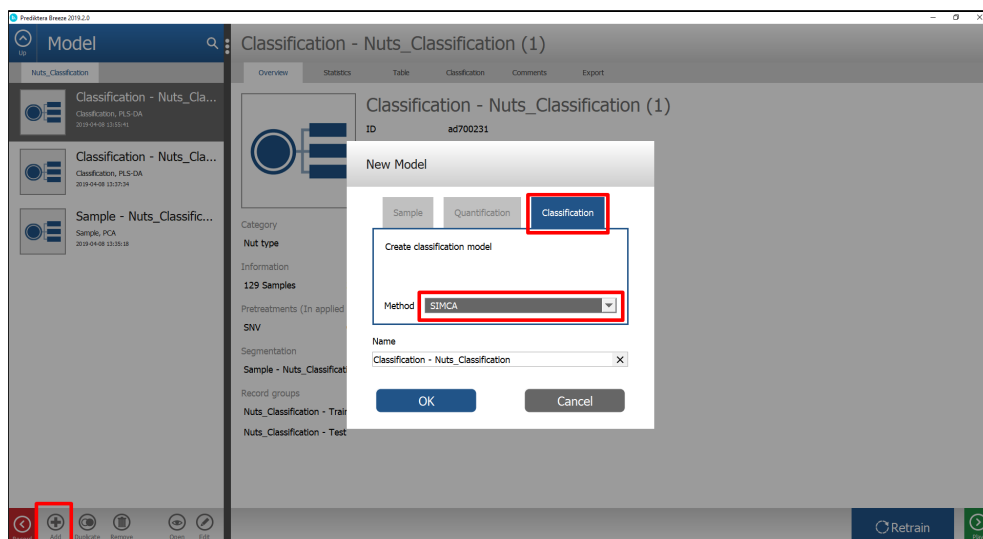
Classes	Total	Almond	Shell	Hazelnut	Pecan	Walnut	No class
Almond	13 (10.1%)	7 (53.8%)		3 (23.1%)			3 (23.1%)
Shell	64 (49.6%)		64 (100%)				
Hazelnut	16 (12.4%)			11 (68.8%)			5 (31.2%)
Pecan	26 (20.2%)				25 (96.2%)		1 (3.8%)
Walnut	10 (7.75%)					10 (100%)	
# Predicted	129 (100%)	7 (5.43%)	64 (49.6%)	14 (10.9%)	25 (19.4%)	10 (7.75%)	9 (6.98%)
Correctly	117 (90.7%)						
Incorrectly	12 (9.3%)						

Type: Train

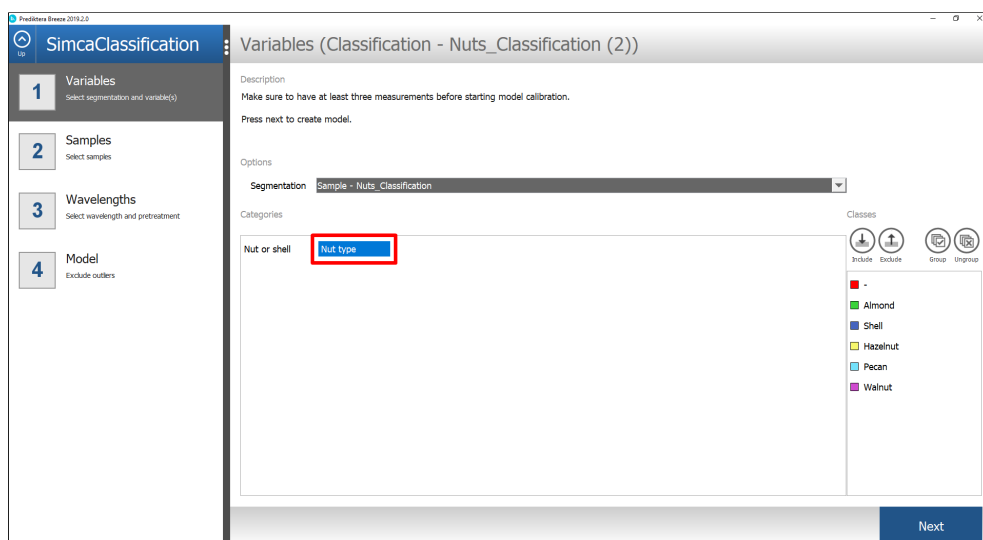
In this example, 3 of the “Almond” samples are misclassified as a “Hazelnut” and 3 as “No class”. For the “Hazelnut” and “Pecan”, there are 5 and 1 samples respectively that are incorrectly classified as “No class”.

Create classification model (SIMCA)

- Let's compare the PLS-DA model with a different classification model type. **Press “Add”, select “Classification” and “SIMCA” as “Method”**. Write a name for the model or use default and **press OK**.

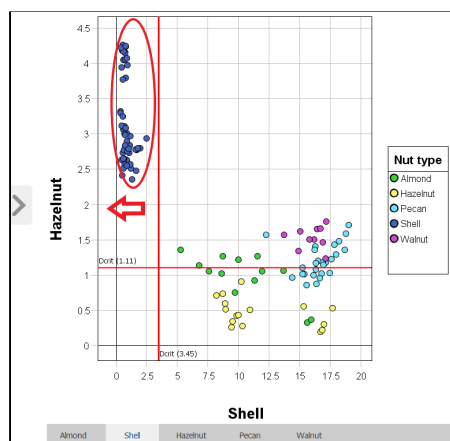
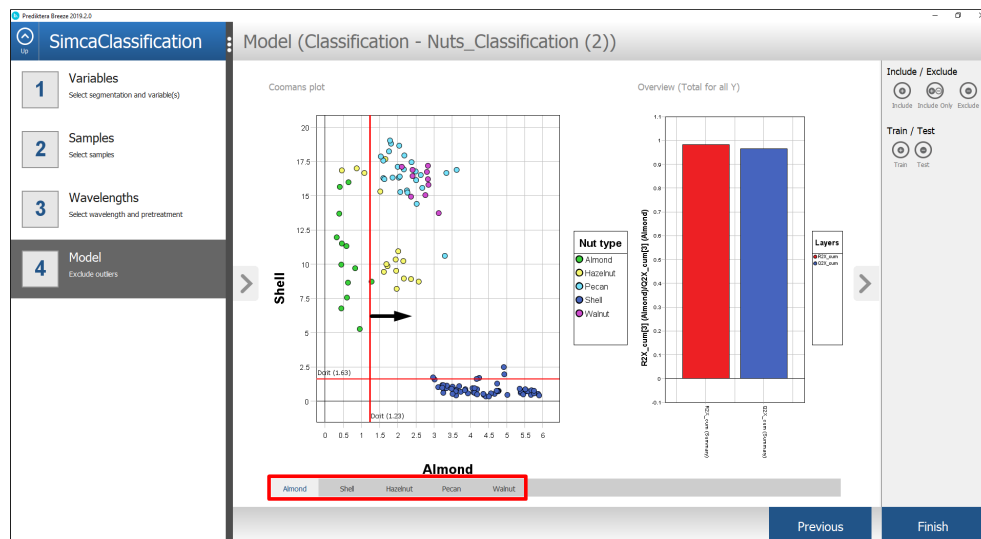


- Choose “Nut type”



- In the 2nd and 3rd step just **press “Next”** (use the default).

4. In the SIMCA method one model is created for each class. All samples are then compared to that class model to determine if they belong to that class. In the Coomans plot you can set the critical distance for each class model. If a sample is inside the critical distance it belongs to the class.
 - a. **Select the class model for “Almond” by using the tabs under the Coomans plot**
 - b. **Drag the vertical red line to adjust the limit to include all “Almond” samples** (but as few as possible of the other samples). Samples to the left of the red line are included in the Almond model.
 - c. **Press the tab for each of the classes and repeat the steps in a. and b.**



Example: All Shell samples (blue dots) are inside the limit of the Shell model

(“Overview (Total for all Y)” is only showing how well each class model can explain

the samples in that class. It does not show how well it can classify other samples)

5. Press “Finish” to complete the model
6. Press the “Classification” tab to see how well the samples in the training data were classified. Press the three dots to maximize the table view and the arrow to open the preview image.



Click on a field in the table to see the corresponding samples in the preview image.

Prediktera Breeze 2019.2.0

Classification - Nuts_Classification (3)

Overview Statistics Table **Classification** Comments Export

Raw B L

Model

Confusion Matrix (For classification models only)

Classes	Total	Almond	Shell	Hazelnut	Pecan	Walnut	No class
Almond	13 (10.1%)	10 (76.9%)		3 (23.1%)			
Shell	64 (49.6%)		64 (100%)				
Hazelnut	16 (12.4%)			16 (100%)			
Pecan	26 (20.2%)			1 (3.85%)	25 (96.2%)		
Walnut	10 (7.75%)					10 (100%)	
# Predicted	129 (100%)	10 (7.75%)	64 (49.6%)	20 (15.5%)	25 (19.4%)	10 (7.75%)	0 (0%)
Correctly	125 (96.9%)						
Incorrectly	4 (3.1%)						

Type Train

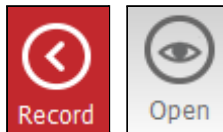
Retrain Play

In this example, with the SIMCA classification a total of only 4 samples were misclassified (you might get slightly different results depending on how you set the critical distance in the previous step) . This can be compared to 12 misclassified samples for the PLS-DA.

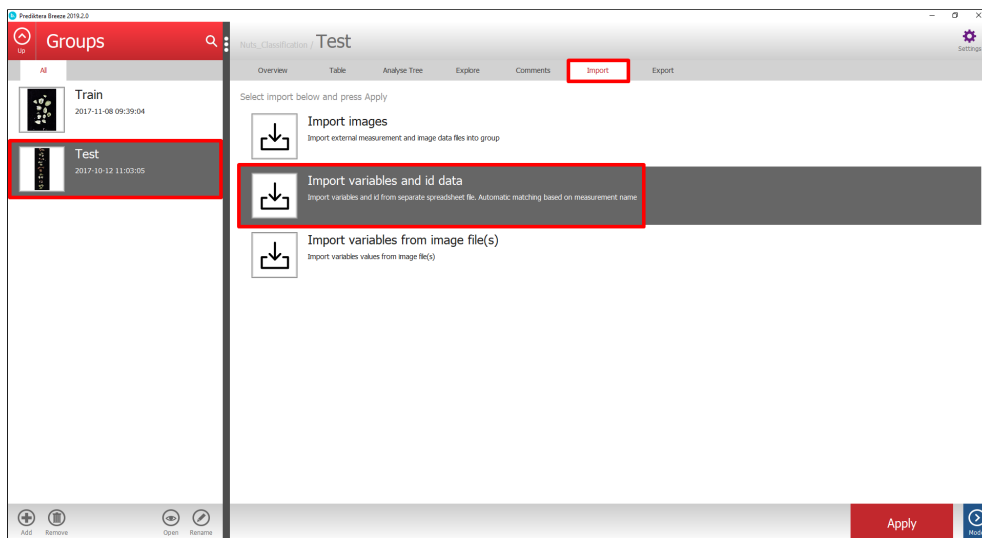
Import the known class information for the test samples

1. To validate the model you should use an external test set to see how well it can classify samples that were not in the training data set. We will now add the known class information to the image “Mix” in the “Test” group.

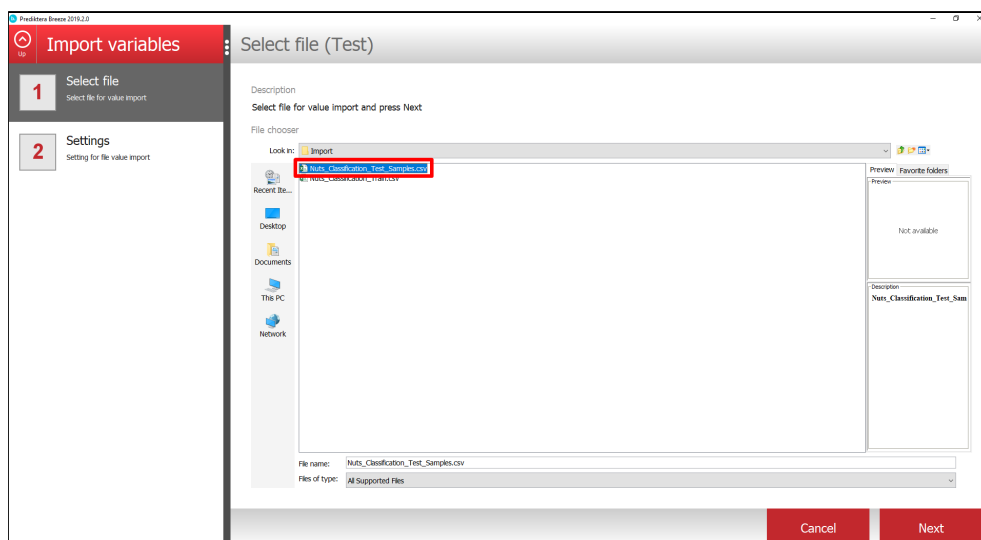
Press the “Record” button and then Press the “Open” button to open the “Nuts_Classification” study.



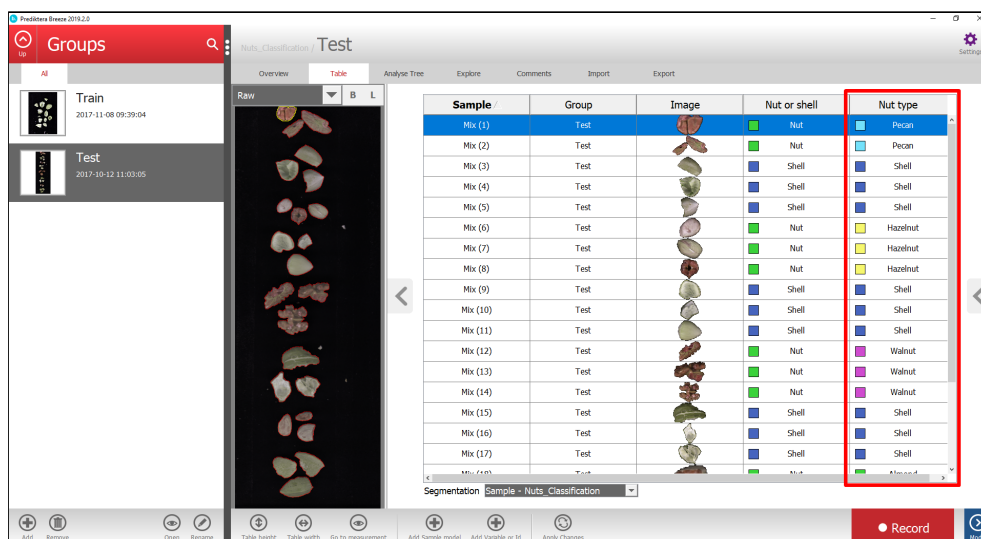
- Select “Test” group
- Select “Import” tab and select “Import variables and id data”
- Press “Apply”.



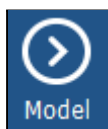
2. Select **"Nuts_Classification_Test_Samples.csv"** and press **"Next"**.

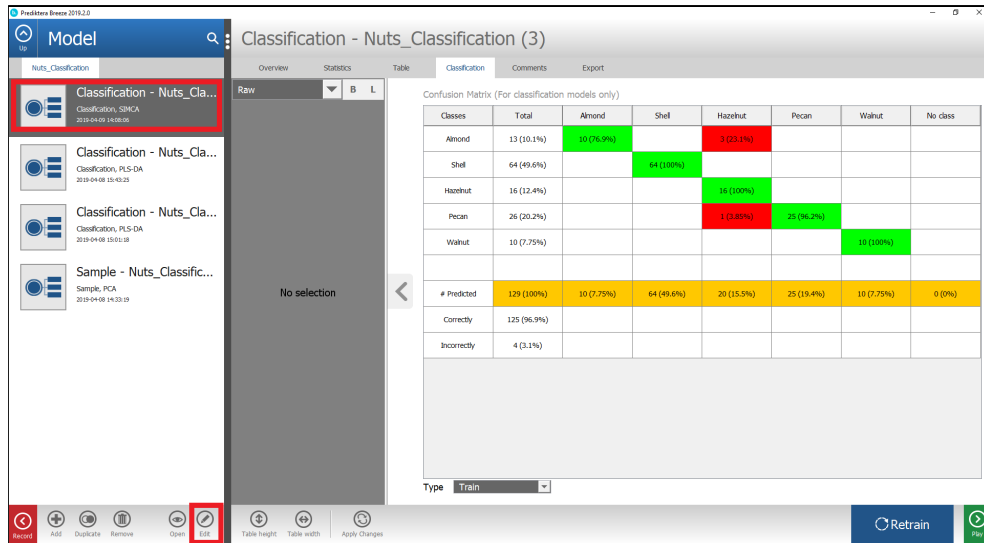


3. In the next step check the option to **"Automatically create Categories"** and press **"Finish"**.
4. The table should look like this for the Test group. Scroll down to see the classes for the **"Mix"** samples.

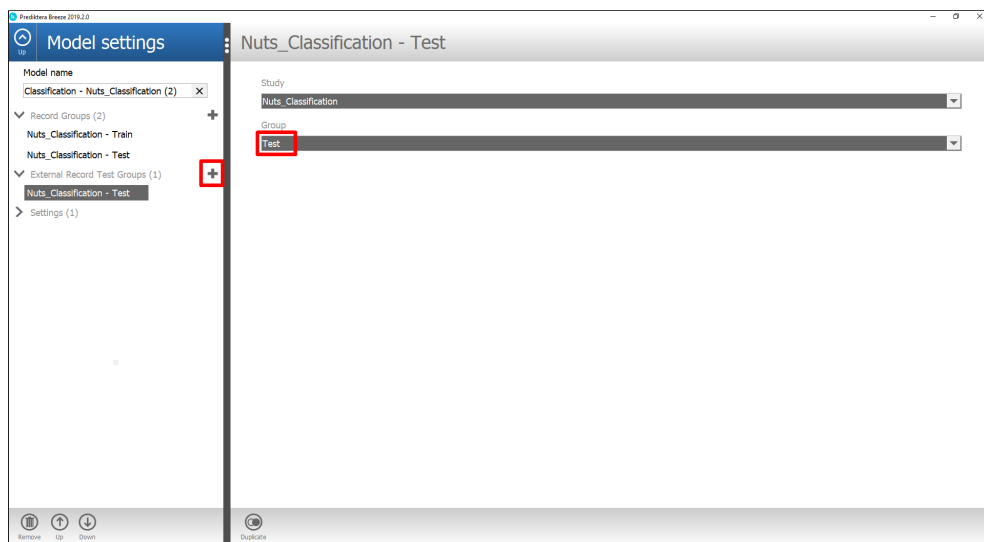


5. Press the **"Model"** button, and with the SIMCA model selected press **"Edit"**





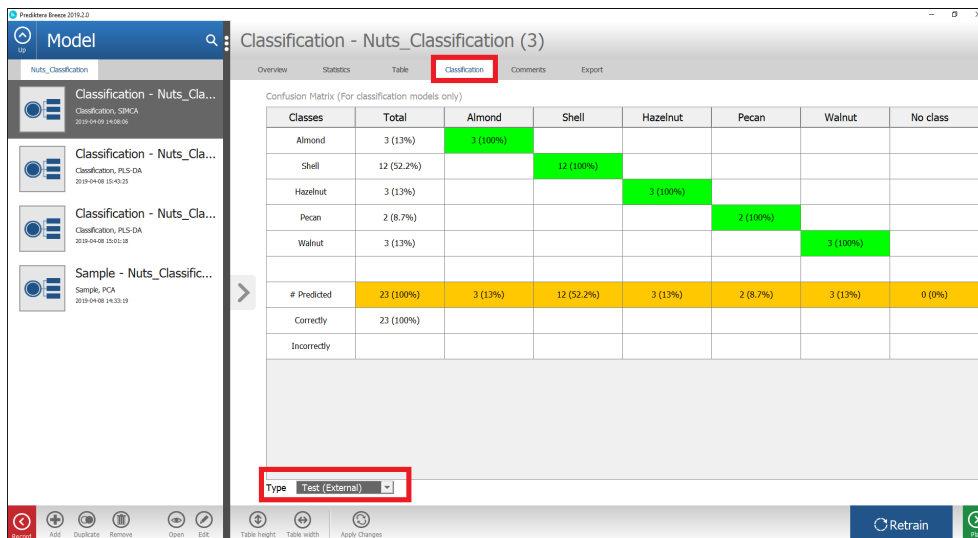
- Press the plus sign to Add “External Record Test Group”. Make sure the group named “Test” is selected in the menu on the right.



- Press the “Up” button located in the upper left corner



- Press the “Classification” tab and make sure that “Test (External)” is selected in the drop down menu under the table.



Classification - Nuts_Classification (3)

Confusion Matrix (For classification models only)

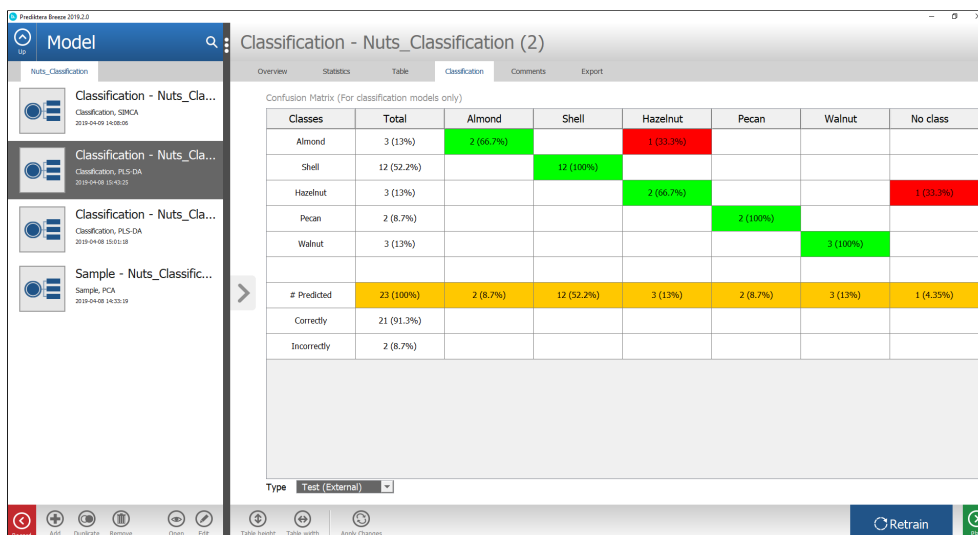
Classes	Total	Almond	Shell	Hazelnut	Pecan	Walnut	No class
Almond	3 (13%)	3 (100%)					
Shell	12 (52.2%)		12 (100%)				
Hazelnut	3 (13%)			3 (100%)			
Pecan	2 (8.7%)				2 (100%)		
Walnut	3 (13%)					3 (100%)	
# Predicted	23 (100%)	3 (13%)	12 (52.2%)	3 (13%)	2 (8.7%)	3 (13%)	0 (0%)
Correctly	23 (100%)						
Incorrectly							

Type: Test (External)

You can now see how well the SIMCA model could classify the 23 samples in the “Test” group.

- Let’s see how well the PLS-DA model can classify the test set. **Select the PLS-DA model, press Edit and add the “Test” group as the “External Record Test Group”** (just like you did for the SIMCA model)

- Look in the Classification table for the PLS-DA. Which model was best?



Classification - Nuts_Classification (2)

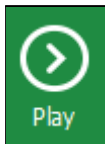
Confusion Matrix (For classification models only)

Classes	Total	Almond	Shell	Hazelnut	Pecan	Walnut	No class
Almond	3 (13%)	2 (66.7%)		1 (33.3%)			
Shell	12 (52.2%)		12 (100%)				
Hazelnut	3 (13%)			2 (66.7%)			1 (33.3%)
Pecan	2 (8.7%)				2 (100%)		
Walnut	3 (13%)					3 (100%)	
# Predicted	23 (100%)	2 (8.7%)	12 (52.2%)	3 (13%)	2 (8.7%)	3 (13%)	1 (4.3%)
Correctly	21 (91.3%)						
Incorrectly	2 (8.7%)						

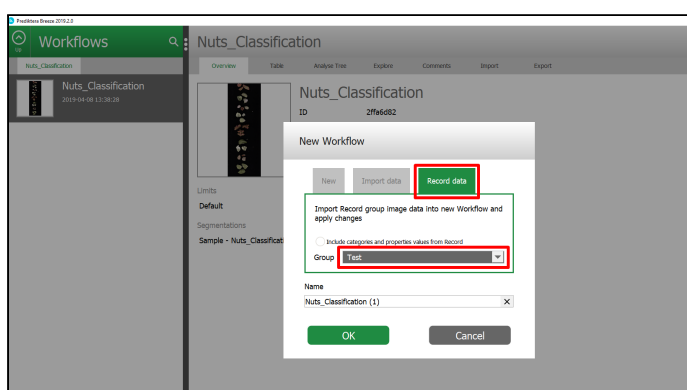
Type: Test (External)

Create workflow and Import Record test data

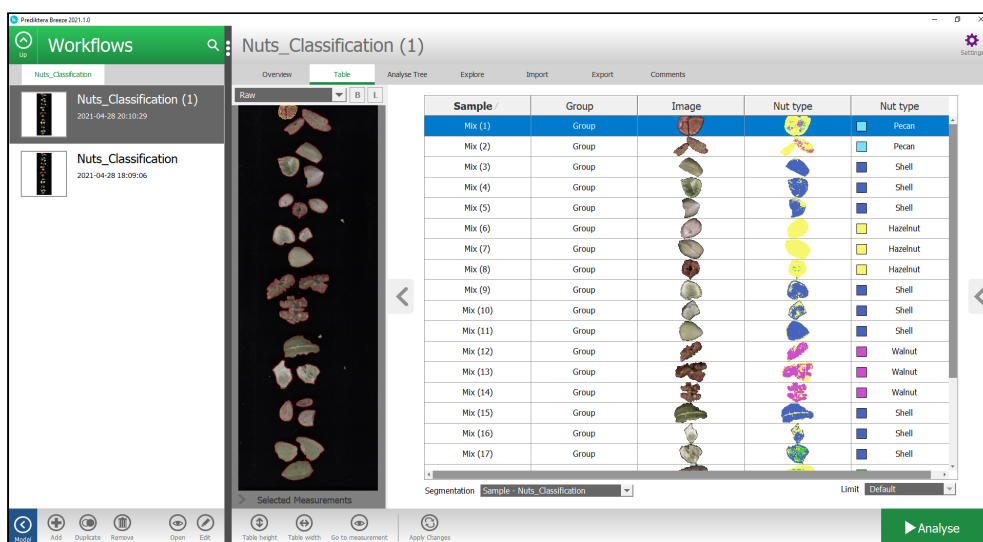
- Now let's test the new models for "Nut type" in the Play mode. **Press "Play"**



- Press the "Add" button, "Record data" tab and select the "Test" group. Press "OK".

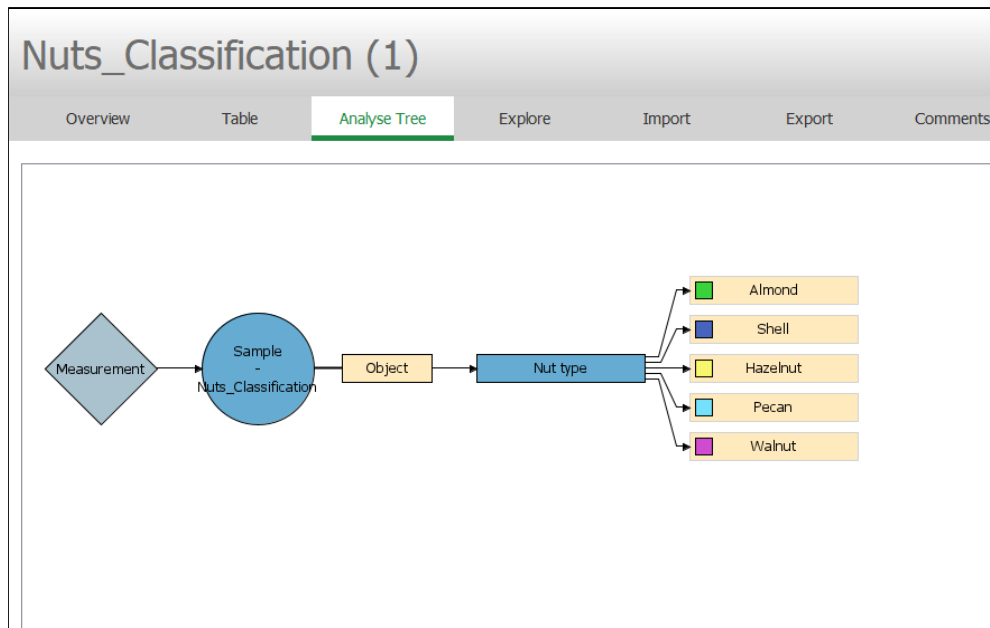


- By default Breeze applies the latest model that you have in "Model" in the workflow. In this case the SIMCA model.

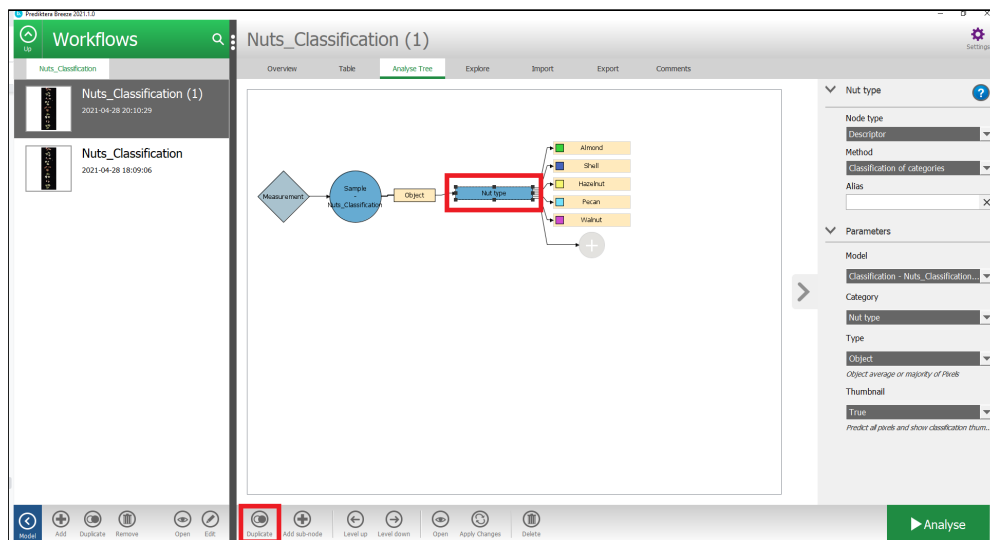


- Press the "Analyse Tree" tab. In the Analysis Tree you can see the steps in the workflow. First the Measurement is analysed by the Sample model that finds the

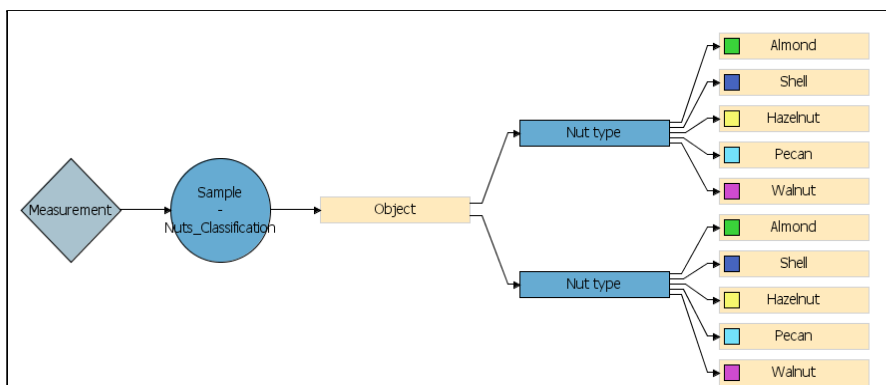
samples (Object). Each Object is then analysed by our model and then classified into different classes.



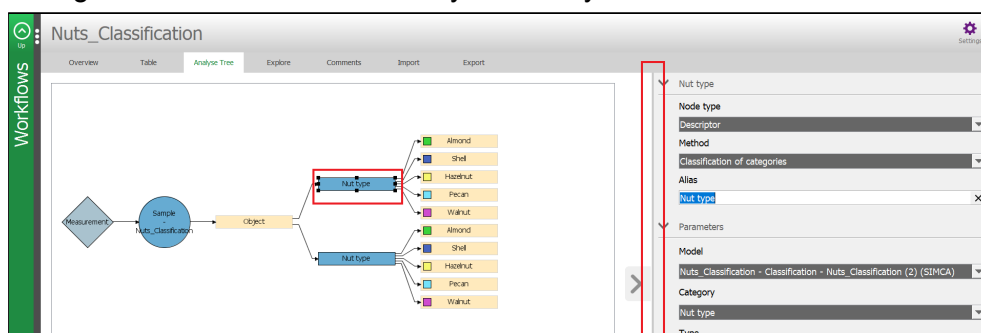
5. Click on the blue symbol for the “Nut type” model (the model that classifies into the Nut or shell classes), and then press the “Duplicate” button.



6. The “Nut Type” model has now been copied and added to the Analyse Tree.

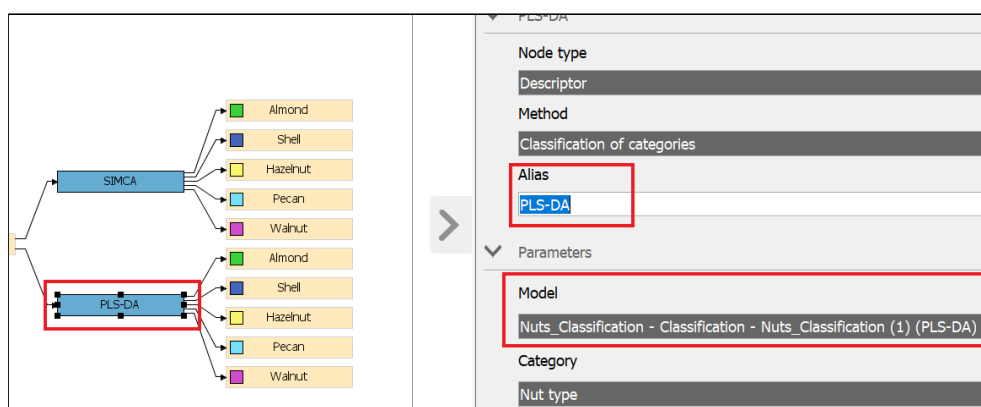


7. **Click on the first “Nut type” model** to see the settings menu for that model on the right side (pull the vertical line to expand the size). In this menu you can see the settings for the selected “Node” in your “Analyse Tree”.

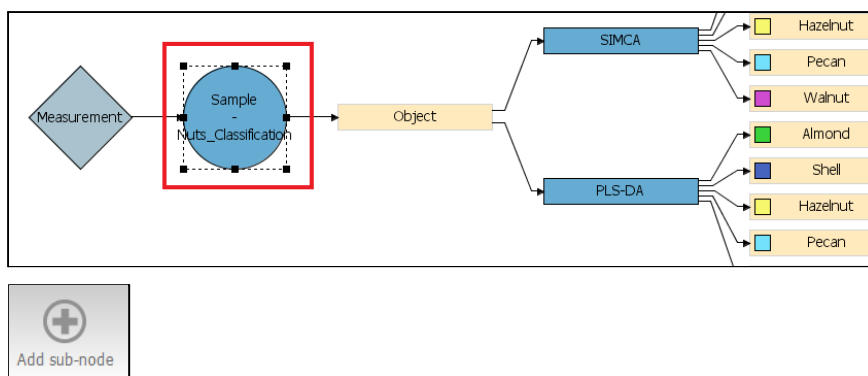


8. In the “Model” drop down menu you can see that this is the SIMCA model. So **in the “Alias” field, write “SIMCA” and press enter on your keyboard.**

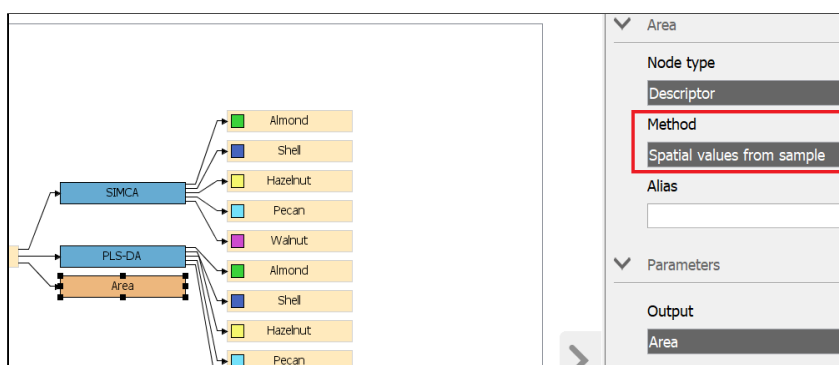
9. **Click on the 2nd “Nut type” model in the Analyse Tree and change to using the PLS-DA model in the “Model” drop down menu. Write the Alias as “PLS-DA” and press enter.** As you can see the text for each model has now been updated in the “Analyse Tree”.



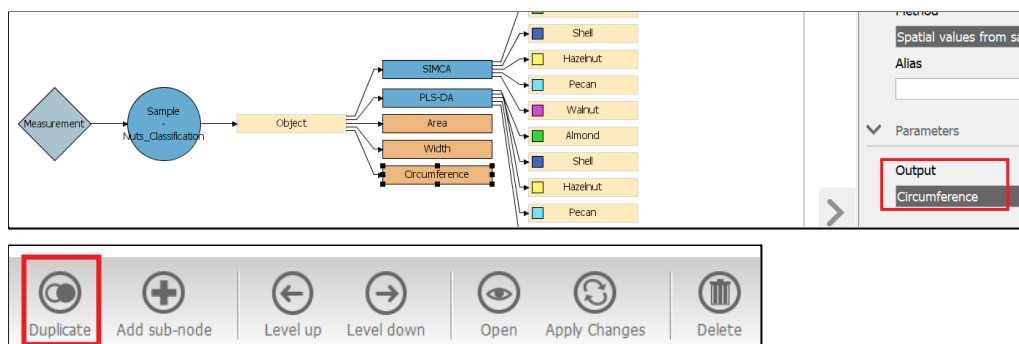
10. In Breeze you can add many different types of descriptors to your workflow. **Click on the sample model in the Analyse Tree and then press “Add sub-node”.**



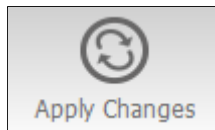
11. A new Node has now been generated as a sub node under the Object generated from the Sample model. In the “Method” drop down menu you can see the different types of descriptors that are available in Breeze. **Select “Spatial values from sample”.** The default “Output” is “area”.



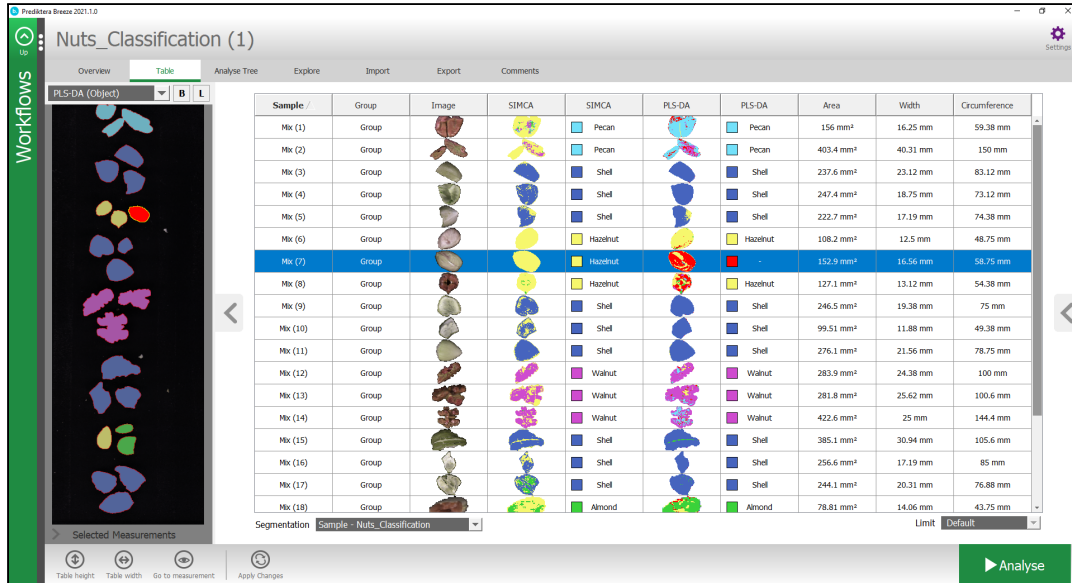
12. **With the Area node selected in the Analyse Tree, press “Duplicate” and select “Width” in the “Output” menu. Press “Duplicate” again and this time select “Circumference”.**



13. Select the Table view and then press “Apply Changes”



14. The “Table” should look like this after it has been updated to the new workflow.



Sample /	Group	Image	SIMCA	SIMCA	PLS-DA	PLS-DA	Area	Width	Circumference
Mix (1)	Group						156 mm²	16.25 mm	59.38 mm
Mix (2)	Group						403.4 mm²	40.31 mm	150 mm
Mix (3)	Group						237.6 mm²	23.12 mm	83.12 mm
Mix (4)	Group						247.4 mm²	18.75 mm	73.12 mm
Mix (5)	Group						222.7 mm²	17.19 mm	74.38 mm
Mix (6)	Group						108.2 mm²	12.5 mm	48.75 mm
Mix (7)	Group						152.9 mm²	16.56 mm	58.75 mm
Mix (8)	Group						127.1 mm²	13.12 mm	54.38 mm
Mix (9)	Group						246.5 mm²	19.38 mm	75 mm
Mix (10)	Group						99.51 mm²	11.88 mm	49.38 mm
Mix (11)	Group						276.1 mm²	21.56 mm	78.75 mm
Mix (12)	Group						283.9 mm²	24.38 mm	100 mm
Mix (13)	Group						281.8 mm²	25.62 mm	100.6 mm
Mix (14)	Group						422.6 mm²	25 mm	144.4 mm
Mix (15)	Group						385.1 mm²	30.94 mm	105.6 mm
Mix (16)	Group						256.6 mm²	17.19 mm	85 mm
Mix (17)	Group						244.1 mm²	20.31 mm	76.88 mm
Mix (18)	Group						78.81 mm²	14.06 mm	43.75 mm

Nice job! You have reached the end of the step 2 of the “Classification of Nuts” tutorial. If you would like to learn more about segmentations and descriptors please try the guides below:

["Segmentation Guide"](#) and ["Descriptor Guide"](#)

or see all tutorial and guides at:

[Breeze - Tutorials and documentation](#)