EMITO-METRIX APPLICATION: QUICK START TUTORIAL

EMito-Metrix is a high-performance pipeline for analyzing mitochondrial morphology and the ultrastructure in multiple tissues from low to high-resolution images acquired with Electron microscopy (EM). Interface allows to compute a set of morphometrics and texture measurements, and provides a list of graphs for optimizing data visualization using dimensionality reduction (UMAP or PCA) and more conventional depiction of data distribution (density curves, histograms, violin plots or star plots). Additionally, a machine learning (ML) module with predictive analytic tools allows determining how any given experimental condition would impact on mitochondrial morphology and ultrastructure

EMito-Metrix plugin was written by Mathieu Vigneau, Emmanuel Doumard and Jean-Philippe Pradère from the RESTORE Institute.

This section describes EMito-Metrix instructions for running and output descriptions

- TABLE OF CONTENTS

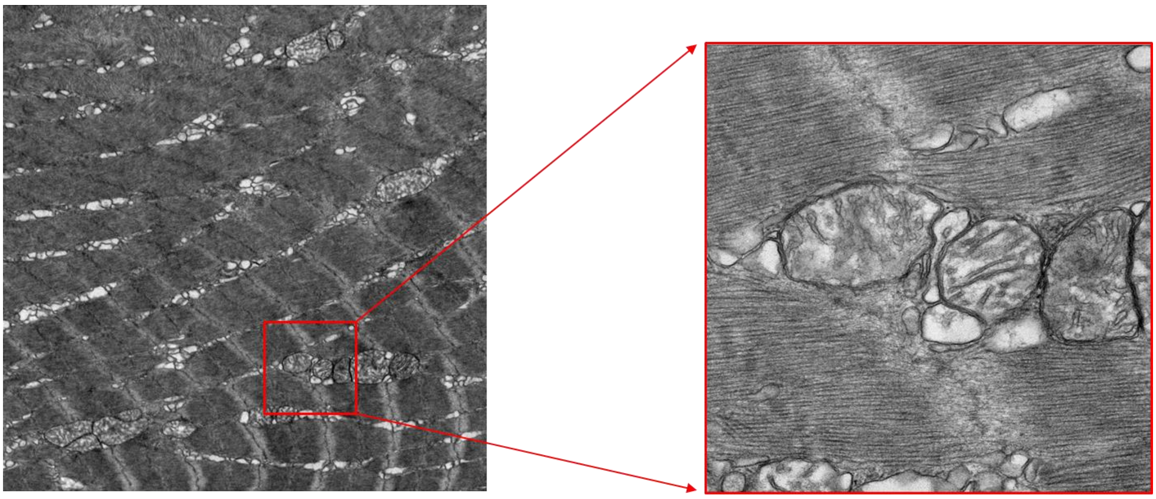
WARNINGS, ADVICES & PREREQUISITES

1- WARNINGS & ADVICES ABOUT ELECTRON MICROSCOPY IMAGES

- About EM magnification & EM image resolution

In electron microscopy (EM) mitochondrial analysis, panoramic images enable the visualization of the tissue but not the details of each mitochondrion. Larger images are useful for gaining finer mitochondria measurements, such as Cristae's orientation, parallelism, and Cristae's quantity within the mitochondria.

In optical microscopy, the resolution of images are closely related to the magnification used during acquisition. The higher magnification, the greater resolution for morphology and ultrastructure details (see figure below).

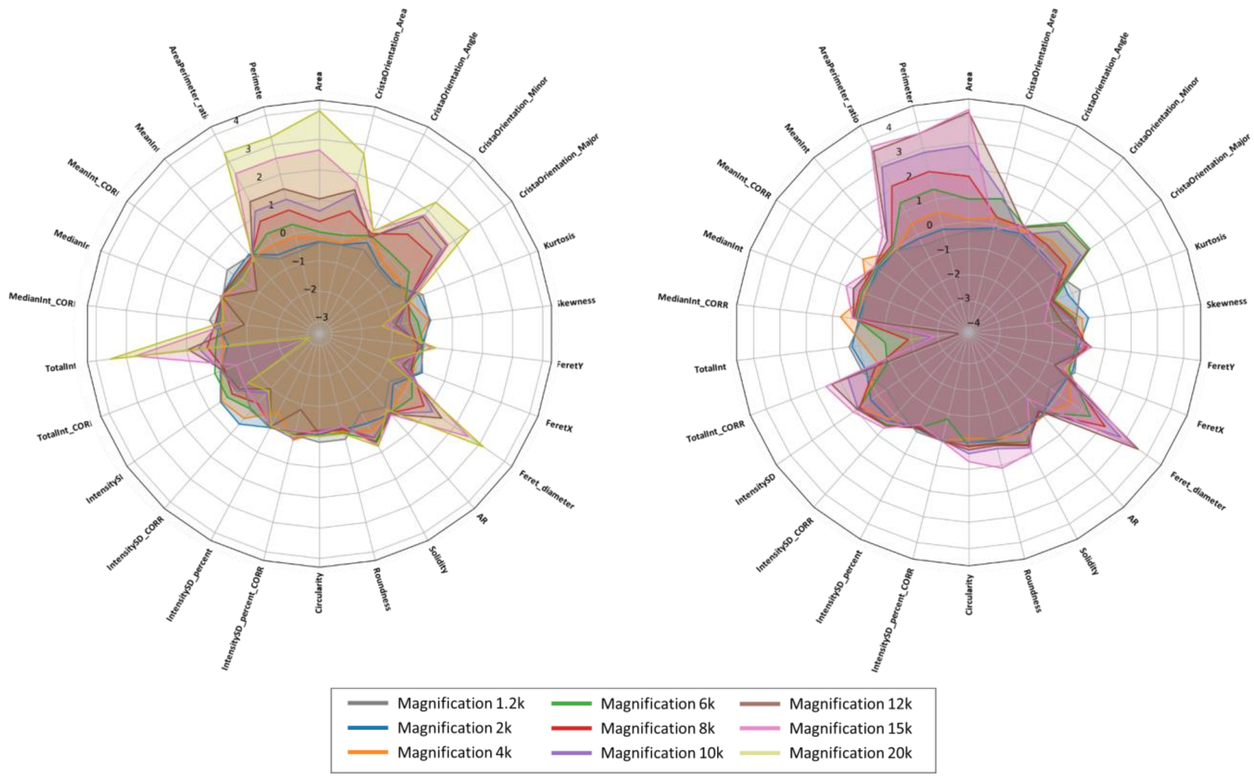


Magnification 2k   Magnification 12k

EM image of the same skeletal muscle biopsy of Zebrafish, using either a magnification of 2k (left panel) or 12k (right panel). The higher magnification, the greater resolution of mitochondrion structure.

Before acquiring your EM images, we recommend that you adjust your magnification to the accuracy you would like to obtain for your morphological and ultrastructure mitochondrial measurements.

Modifying magnification during the acquisition will affect the resolution of EM images, i.e. the resolution of mitochondrial features. Therefore, when running EMito-Metrix analysis, this magnification variation will result in variations of some mitochondrial measurements – especially morphological measurements -, as shown in the following figure.

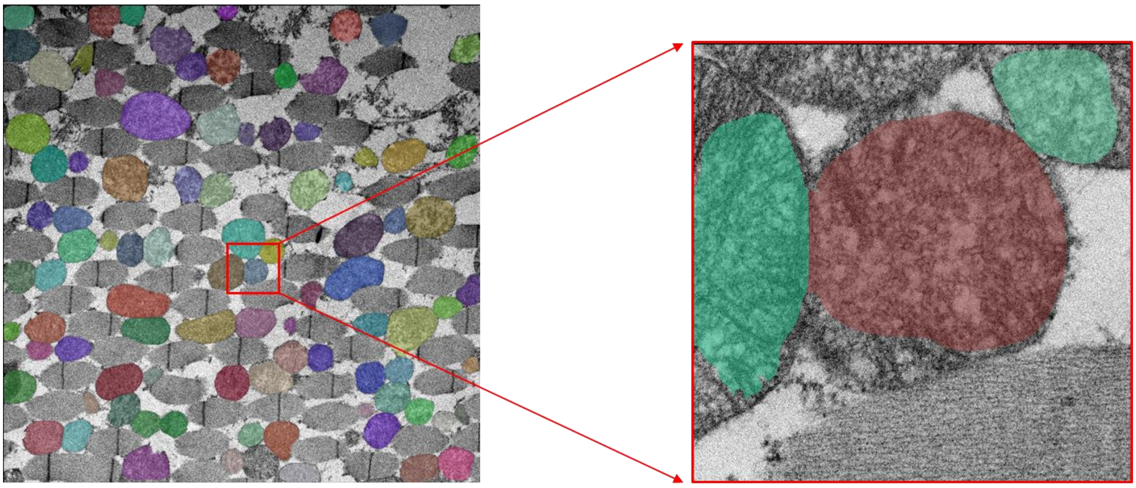


Impact of EM magnification on mitochondria metrics measurements from images of Mouse's skeletal muscles (left panel) and Drosophila's skeletal muscle (right panel)

If you plan to analyze and compare one-by-one several conditions, we highly recommend that you use the same magnification so that you can compare the corresponding morphological measurements

- **About the effect of magnification on our trained model accuracy**

While using a high magnification may be necessary to obtain high-resolution images of mitochondria, it is worth noting that precision and sensitivity of the model used for mitochondria segmentation is highly dependent on mitochondria's environment, i.e. the type of tissue imaged (see below).



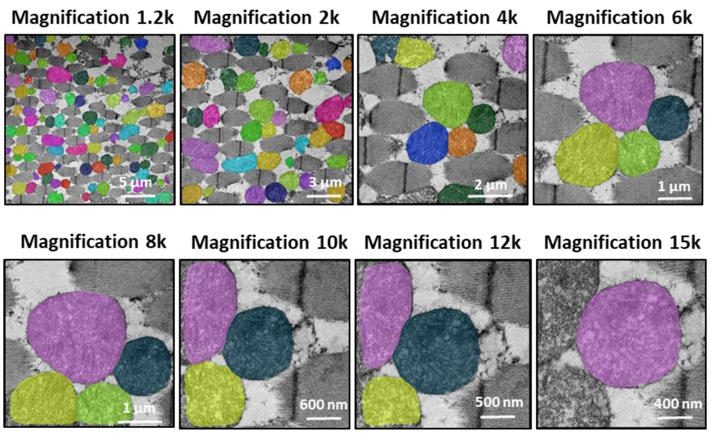
Magnification 1,2k Magnification 15k

EM image of the same skeletal muscle biopsy of Drosophila, acquired at magnification of 1.2k (left panel) or 15k (right panel). For each case, we have projected mitochondria segmentation on the raw EM image. We observe an erroneous segmentation from the 15k magnification, which is due to the absence of tissue's context

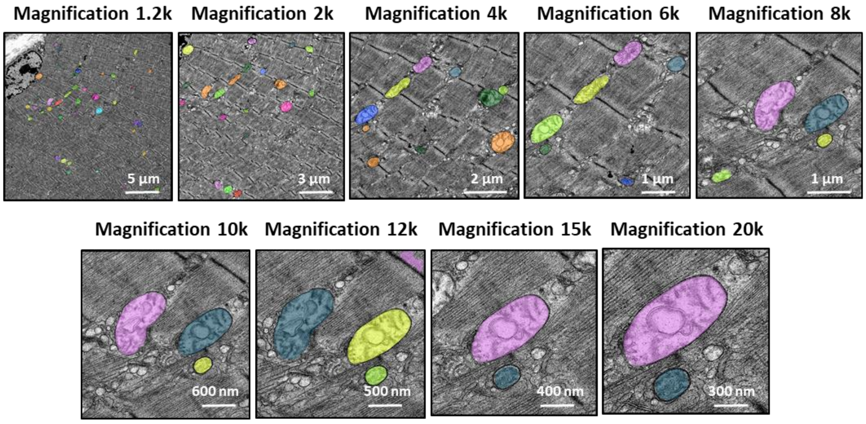
When setting both magnification and tissue's region to acquire, we recommended that you choose an area containing several mitochondria (at least 2 mitochondria) associated with their tissue environment. This should guarantee better mitochondria detection (see below).

We tested if our fine-tuned GM model was working at high magnification. Specie per Specie, we compared mitochondria segmentation from 1,2k to 20k resolution (see figures below). We found that mitochondria segmentation was working very well but with a variable accuracy depending on the specie. Thus, we established that our model works at maximum 15k images for Fly, 12k images for Z-Fish and 20k images for Mouse.

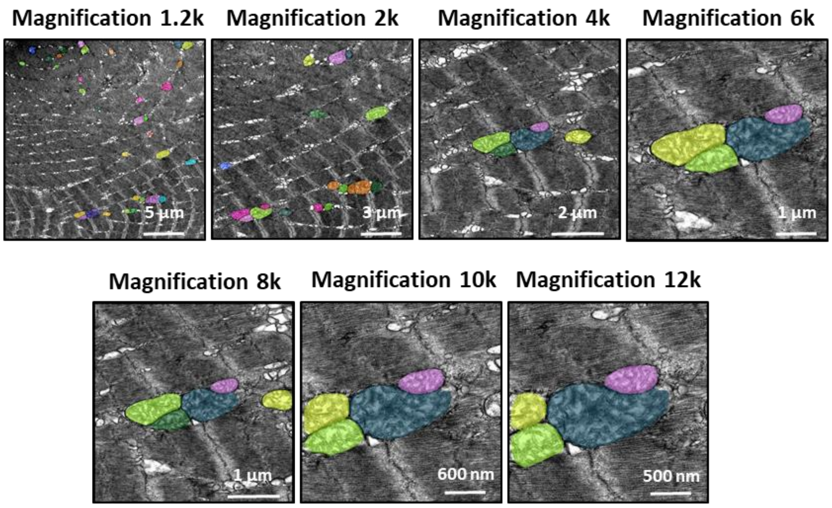
- **Impact of resolution on DROSOPHILA mitochondria segmentation**



- **Impact of resolution on MOUSE mitochondria segmentation**



- **Impact of resolution on ZEBRAFISH mitochondria segmentation**

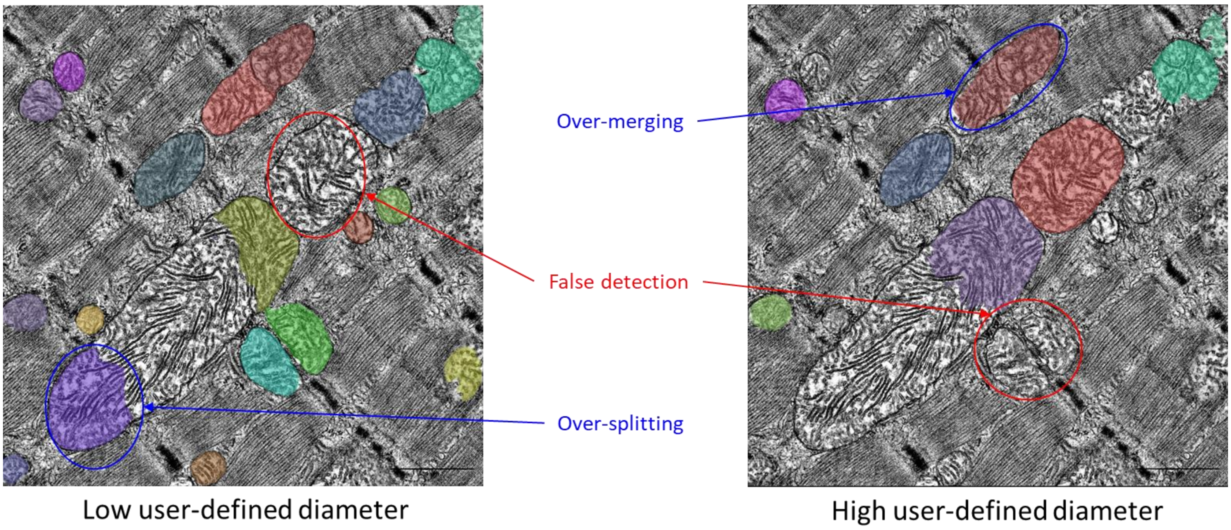


- About the heterogeneity of mitochondria’s size in an EM image

The Cellpose neural network used in our EMitoMetrix application has been separately trained on images of skeletal muscles from different species, resulting in so many different species-specific models. For each model, we annotated mitochondria with the same diameter; however, using Cellpose neural network in an automatic way needs a user-defined mitochondria diameter (in pixel) as an input for segmentation.

User defines the size value as the average diameter of mitochondria from a single image, or from an image-by-image basis. Changing the diameter will change the results of the algorithm outputs (i.e. the objects segmented). When the diameter is set smaller than the true size, then the neural network model may over-split mitochondria. Similarly, if the diameter is set too big, then the neural network model may over-merge mitochondria.

User can define a single diameter per image. Therefore, if the tissue's region imaged contains mitochondria with strong size heterogeneity (i.e. large and small mitochondria in the same image), the neural network model may detect mitochondria with a poor accuracy (for example an over-splitting of large mitochondria and/or an over-merging of small mitochondria, as shown below).



Projection of EM image of skeletal muscle biopsy of Mouse with mitochondria segmentation map (color-coded). We used either a low value (left panel) or a high value as a user-defined diameter for mitochondria detection*

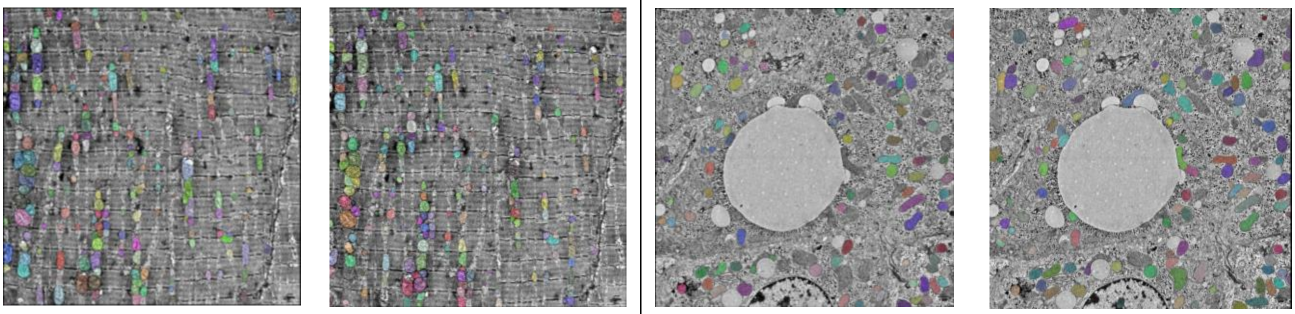
*Images from CBM (CSIC-UAM), UAM University - Laura Formentini.

For better precision and sensitivity of detection, we recommend choosing a tissue's region to acquire with a good homogeneity of mitochondria's size.

- **About the Generalist & Specialist trained models used for mitochondria detection**

As previously explained, each species-specific model has been trained using images of the same tissue - skeletal muscle - from a single specie (i.e. ZebraFish, Mouse, Fly or Human). This implied that dataset used for each model training had a high homogeneity of tissue's environment. On this opposite, we used images of skeletal muscle from several species (i.e. ZebraFish, Mouse, Fly and Human) to train our Generalist model, that is dataset with a very high heterogeneity of tissue's environment. Because of this heterogeneity, and depending on the tissue and/or specie you use, you may have a better mitochondria detection using the Generalist model than the Species-specific models (see below).

Mouse Skeletal muscle | **Mouse Liver**



***Generalist model*: 201 objects** | ***Mice-Specific model*: 230 objects** | ***Mice-Specific model* : 96 objects** | ***Generalist model* : 117 objects**

Projection of EM image of skeletal muscle biopsy (left panels) and liver biopsy (right panels) of Mice with mitochondria segmentation map (color-coded). For each tissue, we used either Generalist or Mice-specific model as an input for Cellpose mitochondria detection. The number of detection is indicated for each condition*

* Images from CBM (CSIC-UAM), UAM University - Laura Formentini

For better precision and sensitivity of detection, we recommend testing both Generalist and Specific models on your dataset, before starting EMitoMetrix application

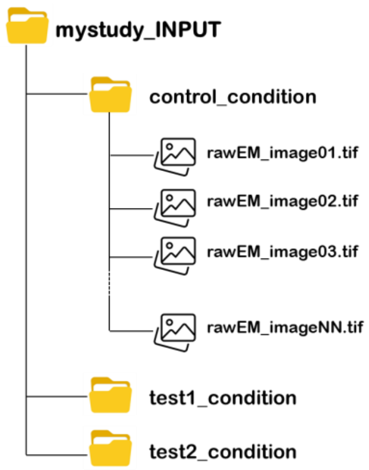
We do not recommend using Cellpose trained models (cyto2, nucleus, cyto), as these models have not been trained on EM images.

2- APPLICATION EXPECTATIONS

- **Expected input files & folders**

EMito-Metrix application allows: 1- analyzing mitochondrial morphology and ultrastructure in EM images **from multiple experimental conditions**; 2- **comparing these conditions one-by-one** in a single (and same) analysis.

Before running such analyze, we recommend that you organize your experimental conditions and input files (i.e. raw EM images) as shown below:



- Save all input files and conditions to analyze and compare in a unique parent folder (*mystudy_INPUT)*
  No preferences regarding the location of mystudy_INPUT folder

- *mystudy_INPUT* must contain as many folders as conditions to compared/analyze
  Example: "control_condition vs test_condition

- *mystudy_INPUT* must contain nothing but the condition folders to compared/analyze.
  No other files or folders saved in mystudy_INPUT folder

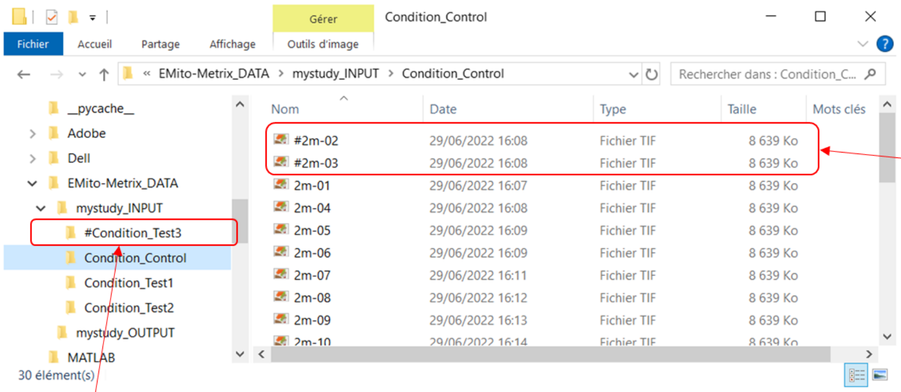
- Each condition folder must contain all input (raw) EM images to analyze.
  Condition folders <u>must contain nothing but the input (raw) images</u>

- Folders name and input files name must respect the following rules:
  no special characters: []!#$%&'()/*,.:;<=>?@^`{|}~]
  no spaces, no accents
  no dot (except the one for file extension)

- We recommend the following file formats for the analysis:
  tif images, tiff images, bmp images

!!Even if accepted, we do not recommend using jpg, jpeg and png file formats, because of their poor resolution!!

To exclude conditions and/or images from the analysis, insert the character # at the beginning of the folder's/image's name you need to exclude (see example below).



Excluded image

Excluded condition/folder

- Expected output folders

Create an output folder (i.e. *mystudy_OUTPUT)* dedicated to output folders and files generated by the EMito-Metrix application. **You must organize *mystudy_OUTPUT* as below:**

- *mystudy_OUTPUT* must be empty when performing the application for the first time

- *mystudy_OUTPUT* name must respect the following rules:
  no special characters: []!#$%&'()/,:;<=>?@^`{|}~]*
  no spaces, no accents

- While running, the application will create output folders and files in *mystudy_OUTPUT* folder.
  mystudy_OUTPUT must contain nothing but outputs.

Do not modify output folders name and images name, which may abort the application running

- Mitochondria segmentation settings

Before running EMito-Metrix application, you must estimate **mitochondria** size as well as **trained model** to use for mitochondria segmentation.

To set these parameters, **we recommend using Cellpose graphical user interface**, following these instructions:

- <u>**Starting Cellpose application**</u>

Open an anaconda prompt

From Windows start menu, type anaconda prompt in the search bar, or open a Terminal window from Linux

To activate your Cellpose python environment, type the following python command line in your opened anaconda prompt, and press Enter:

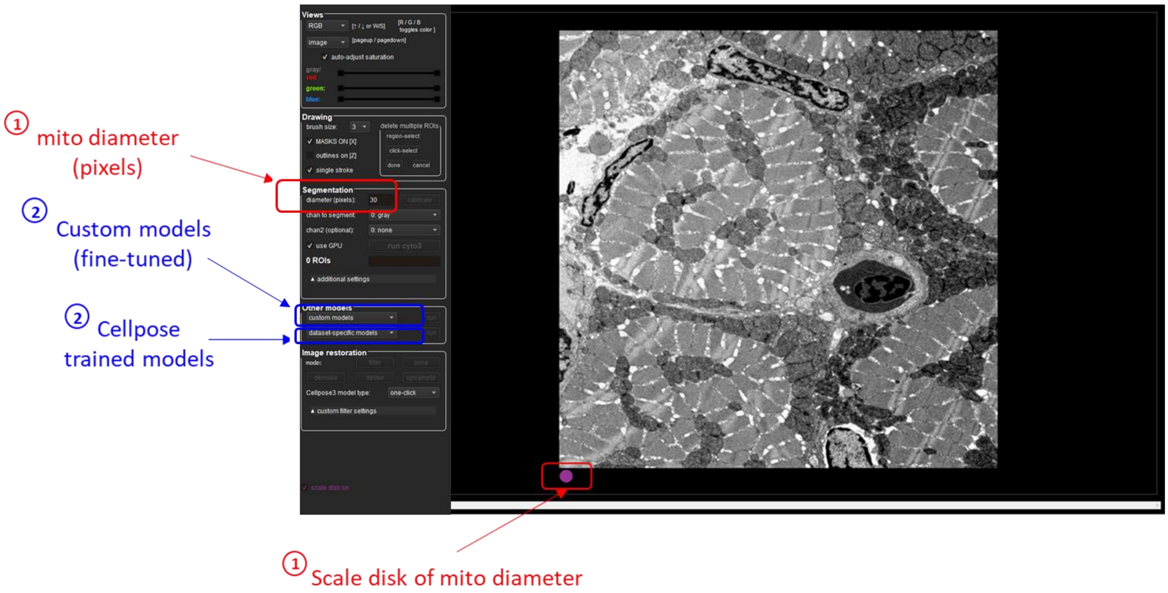
```
conda activate cellpose
```

To start Cellpose graphical user interface, type the following python command lines from your Cellpose python environment and press Enter:

```
cellpose
```

- <u>**Setting mitochondria segmentation**</u>

In Cellpose window, drag and drop your input image to analyze. Here is an example of the Cellpose graphical interface and its functionalities:



Using Cellpose application, estimate the following segmentation settings:

- **Mitochondria size (in pixels):** defined as the average diameter of mitochondria to detect in the image (see [here] for more details about diameter definition)

  Use the purple circle (at the bottom of the view) as a scale disk of the user-defined diameter value

- **Trained model to use for automatic segmentation** (see here for more details)

You can choose either a Cellpose trained model (cyto2, nucleus, cyto) or one of our custom Generalist or Species-specific EM models.

Once you have set both diameter and trained model, click on the *run button* to check the mitochondria segmentation. **Adjust diameter accordingly if the segmentation is not working**.

Depending on mitochondria size heterogeneity of your dataset, you may need to set your diameter value by image, by condition (single value for all images of the condition) or by study (single value for all conditions)

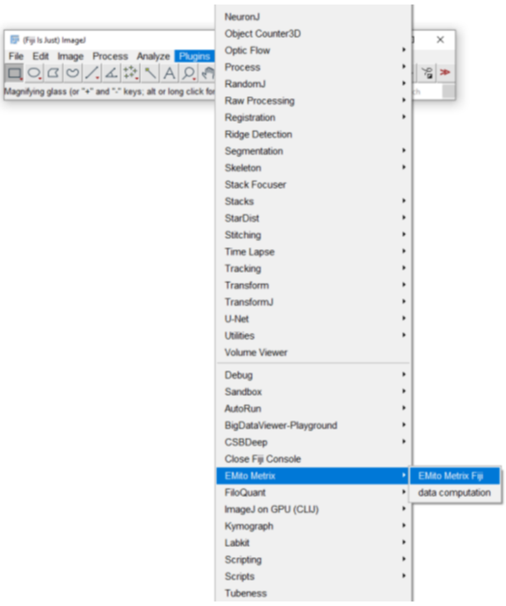
For more details about Cellpose GUI instructions, check out this documentation

EMITO-METRIX RUNNING

1- HOW TO LAUNCH EMITO-METRIX APPLICATION?

EMito-Metrix setting and running are performed using Fiji application (see *EMitoMetrix_Installation* tutorial for installation instructions).

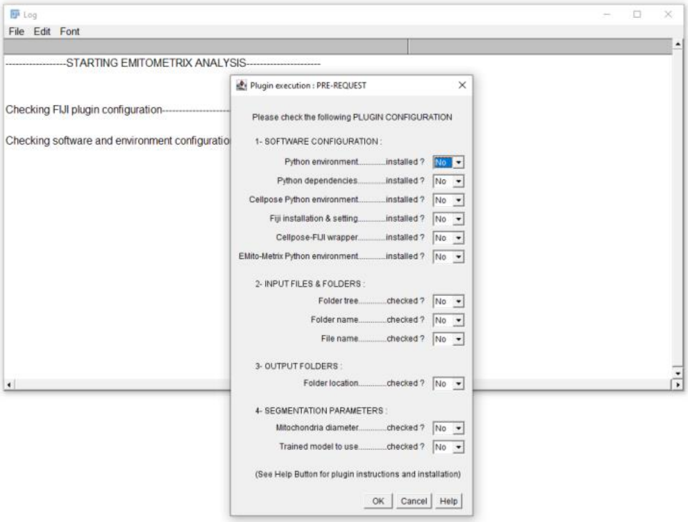
- In your Fiji application directory, double-click the *ImageJ-win64.exe* file to start Fiji software
- Once the Fiji application has started, go to *EMito_Metrix* menu from Fiji *Plugins* menu, and start the application selecting *EMito_Metrix_Fiji,* as shown below :



2- SETTING EMITO-METRIX ANALYSIS

- **Checking EMito-Metrix application & environment**

Once you have launched *EMito-Metrix* application, a window will ask you to confirm that the following application & environment settings are fulfilled:



- **<u>Software configuration</u>** *(see installation guide for detailed instructions)*

  - Python environment
    Python3 environment installed, using Python-like distribution (Anaconda for example)

  - Python dependencies
    Python packages installed for data display and data prediction modules

  - Cellpose Python environment
    Python virtual environment installed and set for Cellpose application. Cellpose GUI installed

  - Fiji installation and setting
    Fiji software distribution installed. Fiji plugin updating and setting ok

  - Cellpose Fiji wrapper
    Cellpose Fiji-wrapper set, using Cellpose python virtual environment

  - Emitometrix Python environment
    Python virtual environment installed and set for data display and data prediction

- **<u>Input files & folders</u>** *(see <u>here</u> for detailed instruction)*

  - Folder tree
    Valid input folders containing condition folders and raw input EM images

  - Folder name
    Valid input folders name and condition folders name

  - File name
    Valid input files name

- **Output folders** *(see [here](#) for detailed instruction)*

  - Folder location

    Valid output folder containing nothing but the output files

- **Segmentation parameters** *(see [here](#) for detailed instruction)*

  - Mitochondria diameter

    Estimate the average mitochondria size (in pixels) for each image/condition to process

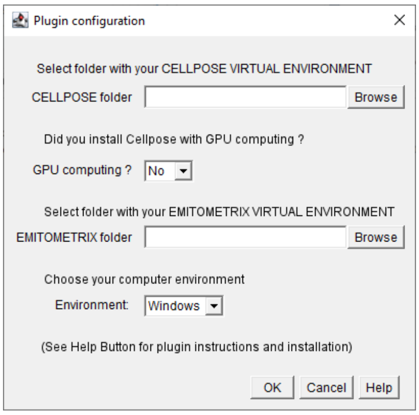
  - Trained model to use

    Define trained model to use for mitochondria detection

If not verified, the application execution will abort and invited you to check each application and environment settings

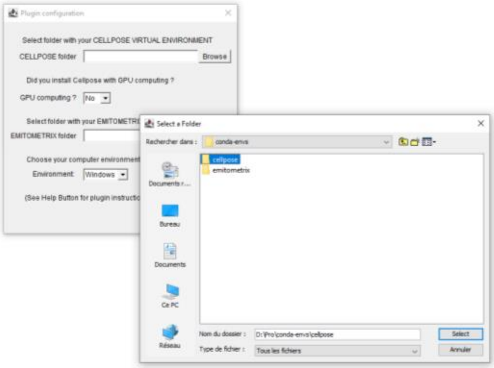
- Setting Python virtual environments

If the EMito-Metrix application installation and setting is correct, the next window will invited you to set these following python parameters:



- **Cellpose folder**: specify parent folder containing your Cellpose virtual environment

  from Windows: C:/Users/username/AppData/Local/anaconda3/envs/cellpose/
  from Linux: /home/username/anaconda3/envs/cellpose/
  from MacOS: /opt/anaconda3/envs/cellpose/



- **GPU computing**: specify if you have install a GPU version of Cellpose

- **EMitometrix folder**: specify parent folder containing your Emitometrix virtual environment
  from Windows: C:/Users/username/AppData/Local/anaconda3/envs/emitometrix/
  from Linux: /home/username/anaconda3/envs/emitometrix /
  from MacOS: /opt/anaconda3/envs/emitometrix /

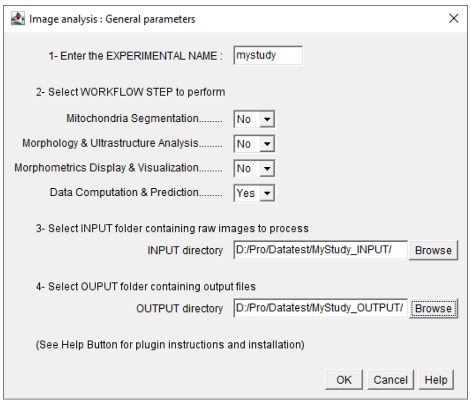
- **Environment**: specify your computer environment used for EMito-Metrix running
  Compatible with Windows, MacOS and Linux environments

The application will check the validity of your specified Cellpose and Emitometrix virtual environments *(see installation instruction guide for Python virtual environment setting)*.

If virtual environment folders are not valid, the application will abort and invite you to check it

- Setting the analysis: workflow steps, input & output folders

If virtual environments are correct, you will have to define the following analysis settings:



- Experimental / protocol name
  String of characters

- <u>**Select Workflow step to perform**</u>

  - Mitochondria Segmentation

    Image preprocessing and mitochondria segmentation

  - Morphology & Ultrastructure Analysis

    Quality control of segmentation & Morphometrics measurement of segmented mitochondria

  - Morphometrics Display

    Data visualization of morphometrics, using graphs & distributions

  - Data computation & prediction

    Predictive analytic tools that determine the impact of one condition on morphometrics

The application checks the following conditions: 1- even if not selected, the "mitochondria segmentation" option is run when the application is unable to find mitochondria segmentation maps in the output folder; 2- even if not selected by user, the application will automatically run "morphology & ultrastructure analysis" option when unable to find mitochondria metrics files in the output folder.

The option to run "morphometrics display" and "data prediction" without selecting "Mitochondria segmentation" and "morphology & ultrastructure" steps allows running separate condition comparisons, without restarting mitochondria detection and morphometrics measurements (less time consuming).

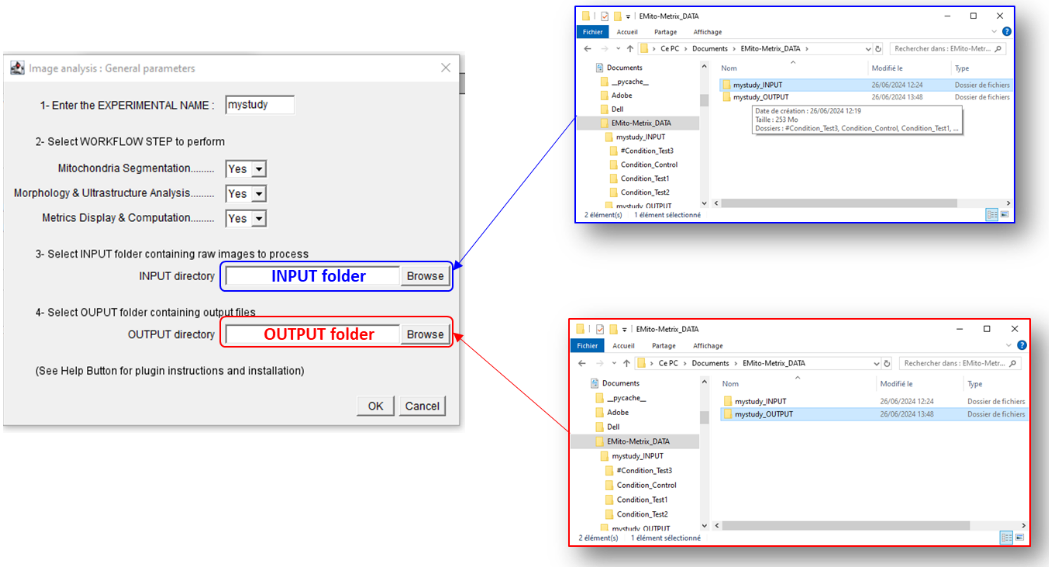
- <u>**Input & output folders**</u>

  - **INPUT Directory**: specify root directory containing condition folders and input files

    Example: C:\Users\username\Documents\EMito-Metrix_DATA\mystudy_INPUT

  - **OUTPUT Directory**: specify root directory used to save output files, images and measurements

    Example: C:\Users\username\Documents\EMito-Metrix_DATA\mystudy_OUTPUT

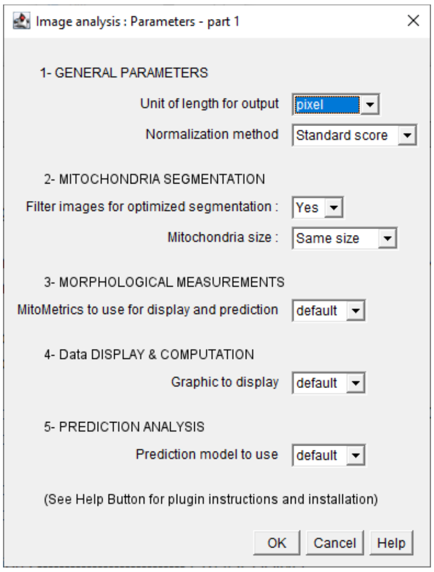


The application will check the validity of INPUT and OUTPUT folders and files *(see here for detailed instruction)*

If INPUT and/or OUTPUT folders and files are not valid, the application execution will abort and invite you to check INPUT and OUTPUT folders/files.

- Setting the analysis: workflow parameters for each step

Once INPUT and OUTPUT folders are valid, define settings for each workflow step previously selected, as shown below:

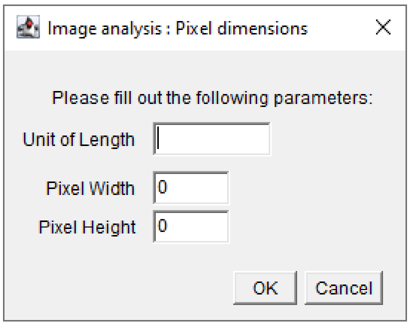


- **General parameters**

  - **Unit of length**: unit used for morphometrics measurements

    pixel or calibrated

  If you select *calibrated*, you will have to define *pixel width*, *pixel height* and *unit of length*.



- **Normalization method**: Raw EM images are normalized on amplitude to attenuate global gray level variations observed between images, which may be related to variations in the amount of staining agent.

<u>min-max scaling</u>: consists in rescaling the range of features to scale the range in [0, 1]. This method is preferred when your data does not follow a normal distribution. Does not handle outliers well.

<u>standard score</u>: values (x) are centered on mean (mean(x)) with a unit standard deviation (sd(x)). This method is preferred when your data follows a normal distribution. Handles outliers well.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Min-max scaling

$$x' = \frac{x - \mathrm{mean}(x)}{\mathrm{sd}(x)}$$

Standard score

- **<u>Mitochondria segmentation</u>**

Filtering images for optimized segmentation: when EM images are acquired with high resolution and/or high magnification, mitochondria detection can be improved using Gaussian filtering

Yes or No (default value)

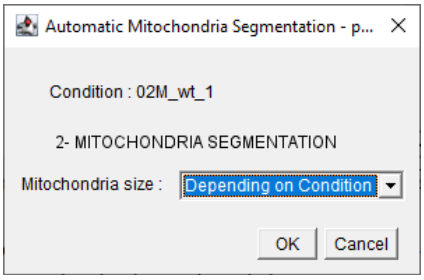
Gaussian filtering applied <u>for segmentation only</u>, not for mitometrics analysis

Mitochondria size: decide if mitochondria size varies across conditions or/and images or is the same for all conditions and images

same size (default value) or variable size (across conditions and/or images)

If you select *variable size*: select the level of mitochondria size variability

"Depending on condition" or "Depending on images"



- **<u>Morphological measurements</u>**

Morphological and texture measurements used for data display and prediction analysis

default (all metrics) or custom (user-defined metrics)

- **Data display & computation**

  Graphs and data distributions used for morphometrics visualization

  default (all data distributions) or custom (user-defined distributions)

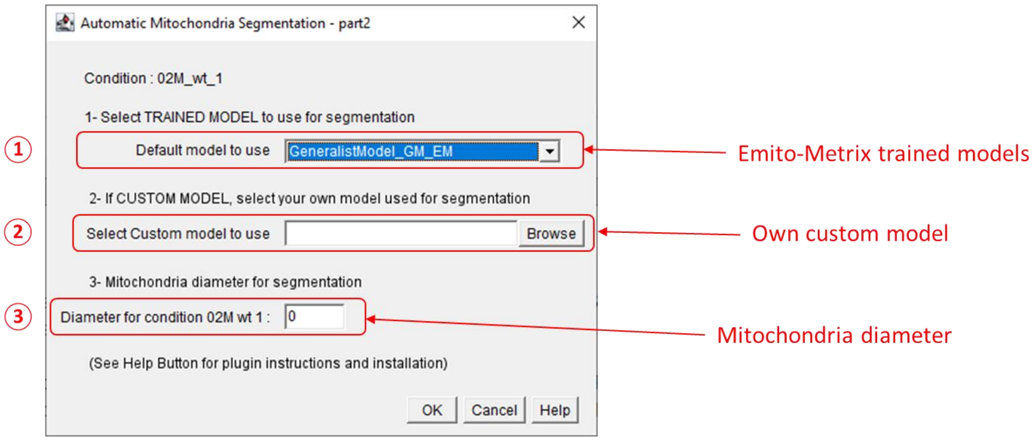
- **Prediction analysis**

  Machine learning models used for data prediction analysis

  default (all models) or custom (user-defined models)

- Setting the analysis: mitochondria segmentation parameters

Depending on the options set in the previous section (see here), you will have to define the following mitochondria segmentation settings:



- Trained model used for mitochondria segmentation (see here for detailed instruction):

Cellpose models (cyto, cyto2 or nucleus), EMito-Metrix trained models (Specialist or Generalist models) or custom models (your own trained/fine-tuned model)

- If custom model: specify your own custom trained/fine-tuned model

Select your trained/fine-tuned model file with a dialog box

- Mitochondria diameter (in pixels): average size of mitochondria to detect in the image (see here for detailed instruction)

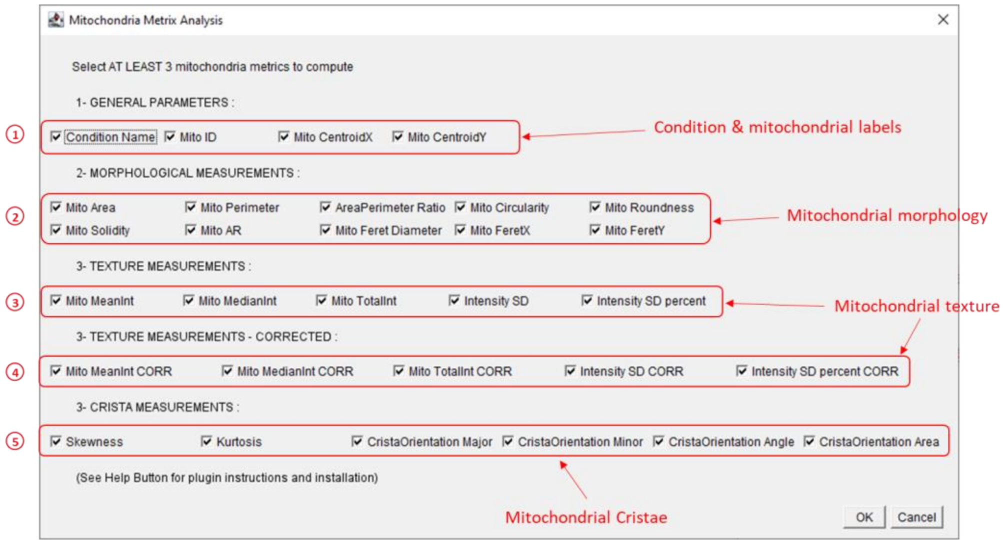
Same size: *the user-defined value must be set only once.*

Variable size, depending on condition*: the user-defined value must be set for each condition*

Variable size, depending on image*: the user-defined value must be set for each image*

- Setting the analysis: morphometrics measurements selection

Next, you will have to select morphological and texture measurements to use for data display and prediction analysis.

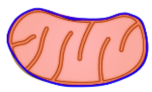
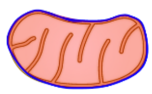


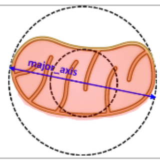
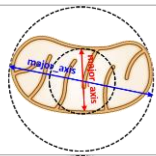
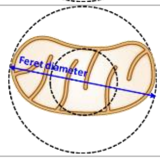
- <u>**General parameters**</u>

  Condition & images name, mitochondria labels and spatial position (Centroid X&Y)

- <u>**Morphological measurement**</u>

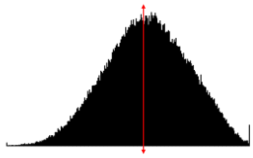
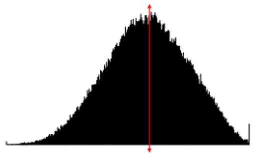
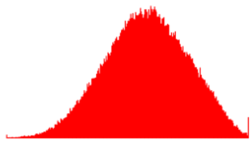
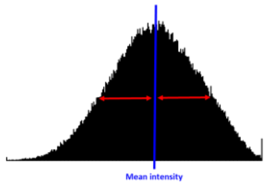
  Mitochondria shape measurements

METRIC NAME	DEFINITION	Illustration
Mito_Area	Area of the mitochondria	
Mito_Perimeter	The length of the outside boundary of the mitochondria.	
AreaPerimeter_Ratio	Ratio between Perimeter and Area	
Mito_CentroidX and Mito_CentroidY	X and Y coordinates of the center point of the mitochondria.	
Mito_Circularity	Shape of the mitochondria defined as ***(4 π *area/perimeter^2)***. A value of 1.0 indicates a perfect circle. As the value approaches 0.0, it indicates an increasingly elongated shape	

Mito_Roundness	Shape of the mitochondria defined as ***(4*area/(π*major_axis^2))***, or the inverse of the aspect ratio.	
Mito_Solidity	Ratio between area and convex area	
Mito_AR	Aspect Ratio : Ratio between major axis and minor axis	
Mito_Feret_Diameter	The longest distance between any two points along the mitochondria boundary, also known as maximum caliper	
Mito_FeretX and Mito_FeretY	Starting coordinates of the Feret's diameter	

- **<u>Texture measurements</u>**

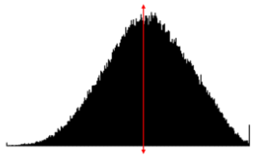
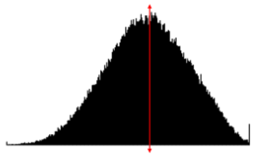
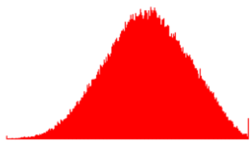
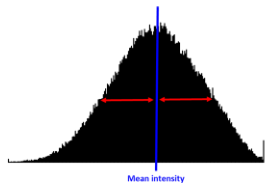
Ultrastructure measurements, based on distribution of normalized gray values in mitochondria

METRIC NAME	DEFINITION	Illustration
Mito_MeanInt	Average intensity calculated from mitochondria's normalized gray values	
Mito_MedianInt	Median intensity calculated from mitochondria's normalized gray values.	
Mito_TotalInt	Sum of the mitochondria's normalized gray values	
Intensity_SD	Standard deviation of the mitochondria's normalized gray values used to generate the mean intensity. Measure of crista's density within the mitochondria	
Intensity_SD_percent	Ratio between Intensity_SD and MeanInt. Measure of crista's density within the mitochondria	

- **Texture measurements – corrected**

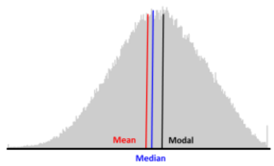
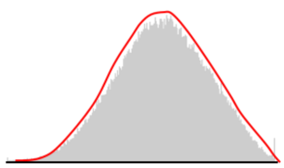
Ultrastructure measurements, based on distribution of corrected gray values calculated in mitochondria.

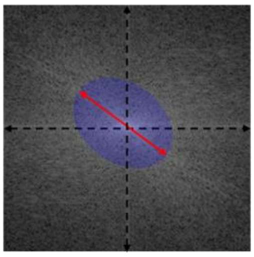
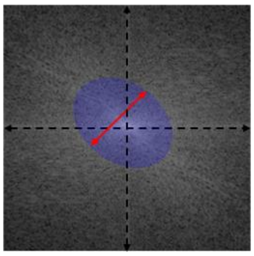
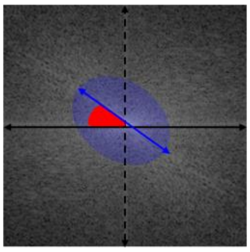
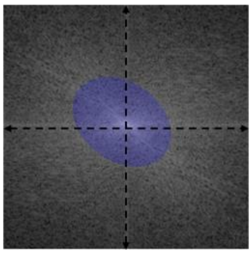
Corrected gray values are calculated from normalized gray values, by filtering noise with a high frequency filter (Fast Fourier Transformation).

METRIC NAME	DEFINITION	Illustration
Mito_MeanInt_CORR	Average intensity calculated from mitochondria's gray values after High frequency filtering (FFT noise correction)	
Mito_MedianInt_CORR	Median intensity calculated from mitochondria's gray values after High frequency filtering (FFT noise correction)	
Mito_TotalInt_CORR	Sum of the mitochondria's gray values, after High frequency filtering (FFT noise correction)	
Intensity_SD_CORR	Standard deviation of the mitochondria's gray values used to generate the mean, after High frequency filtering (FFT noise correction). Measure of crista's density within the mitochondria	
Intensity_SD_percent_CORR	Ratio between Intensity_SD_CORR and MeanInt_CORR. Measure of crista's density within the mitochondria	

- **Crista measurements**

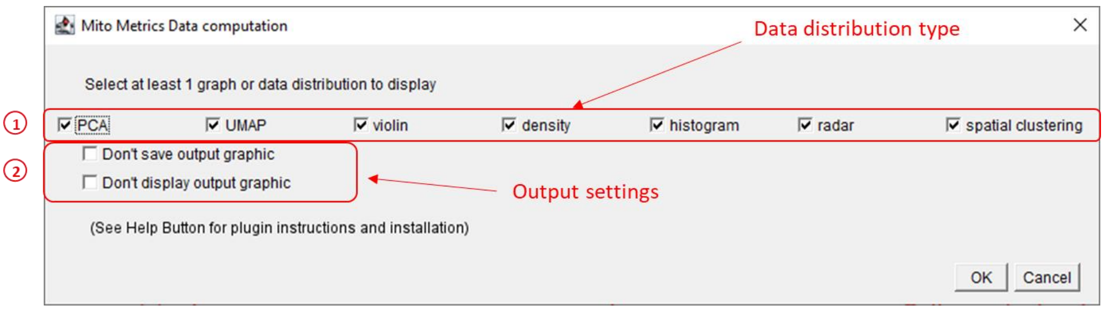
Measurements of cristae's organization and orientation

METRIC NAME	DEFINITION	Illustration
Skewness	Third order moment about the mean. Measure of the asymmetry of the mitochondria's normalized gray values about the mean intensity	
Kurtosis	The fourth order moment about the mean. Measure of the "tailedness" of the mitochondria's normalized gray values about the mean intensity.	

CristaOrientation_Major	Primary axis of the best fitting ellipse calculated from the frequency spectrum of the mitochondria's normalized gray values. Measure of the Crista's orientation, alignment and number within the mitochondria.	
CristaOrientation_Minor	Secondary axis of the best fitting ellipse calculated from the frequency spectrum of the mitochondria's normalized gray values. Measure of the Crista's orientation, alignment and number within the mitochondria.	
CristaOrientation_Angle	Angle (between the primary axis and a **line parallel to the x-axis** of the image) of the best fitting ellipse calculated from the frequency spectrum of the mitochondria's normalized gray values. Measure of the Crista's orientation, alignment and number within the mitochondria.	
CristaOrientation_Area	Area of the best fitting ellipse calculated from the frequency spectrum of the mitochondria's normalized gray values. Measure of the Crista's orientation, alignment and number within the mitochondria.	

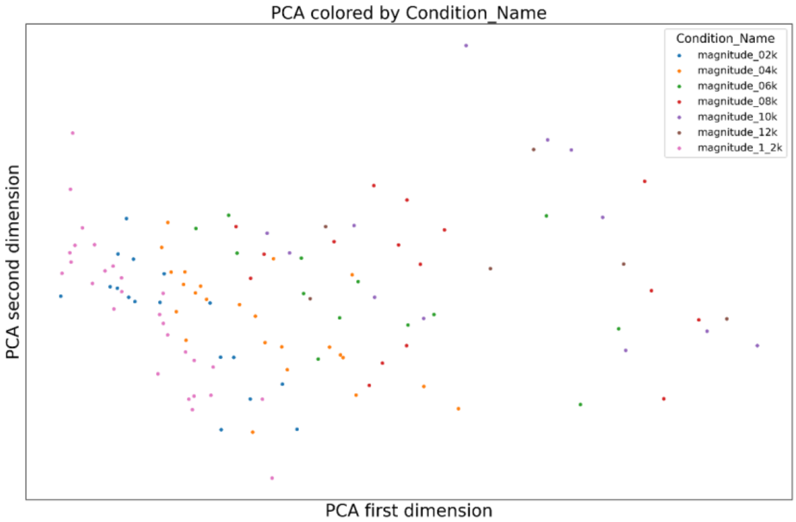
- Setting the analysis: graphs and data distribution selection

Then, select graphs and data distributions used for morphological and texture visualization:



- **PCA distribution**

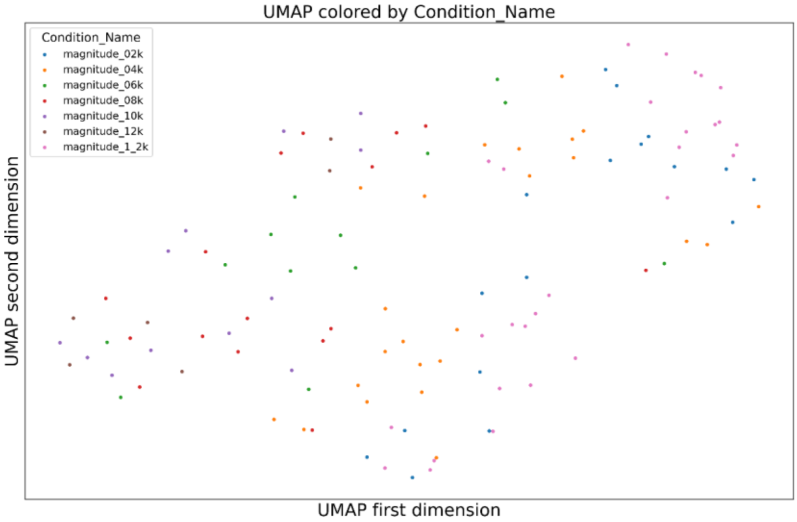
  The principal component analysis (PCA) is a linear dimensionality reduction technique that consists in transforming data onto a new coordinate system such that dimensions are orthogonal and capture the most variation in the data. We keep only the first two dimensions to be plotted.



PCA distribution of mitochondria (one point per mitochondria), according to the magnitude used for EM acquisitions.

- **UMAP distribution**

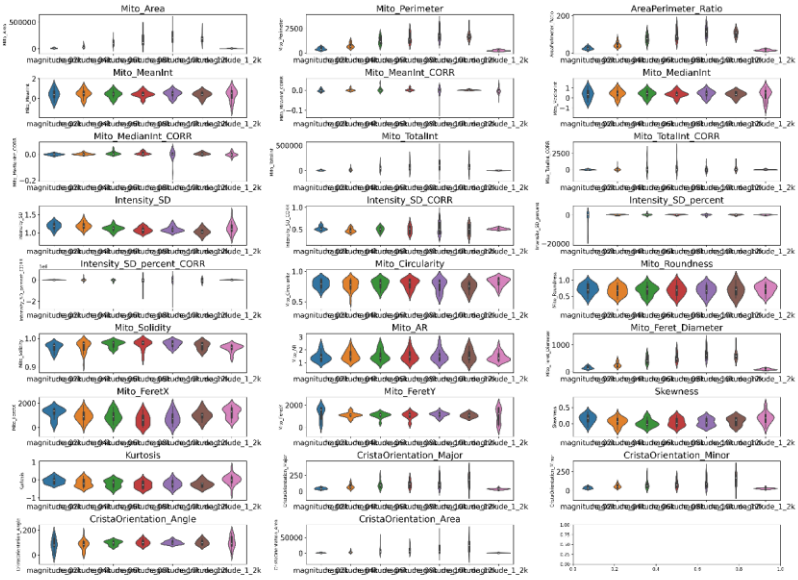
  The Uniform Manifold Approximation for Projection (UMAP) is a non-linear dimensionality technique based on Rimannian manifold. It connects each point to its nearest neighbors in high dimension, and project the manifold in low dimension to be plotted. Its properties ensure that closest on the projection are close in high dimension, and points that are far from each other on the projection are far from each other in high dimension.



UMAP projection of mitochondria (one point per mitochondria), according to the magnitude used for EM acquisitions.

- **<u>Violin distribution</u>**

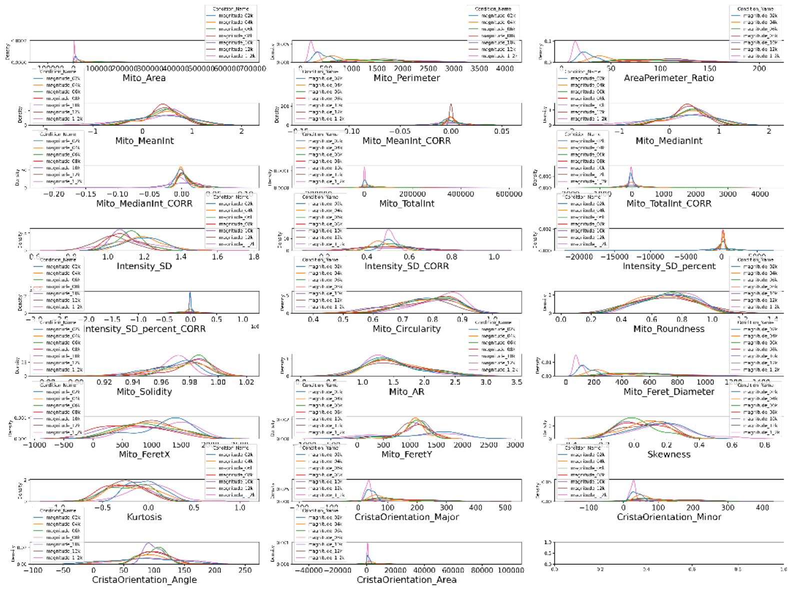
  A violin plot is used to easily compare distributions between conditions. Each condition is plotted on a separate axis, but they are aligned with each other for comparison. The density is smoothed by a kernel density estimator.



Violin plots of mitochondria measurements (one plot per measurement), according to the magnitude used for EM acquisitions.

- **<u>Density distribution</u>**

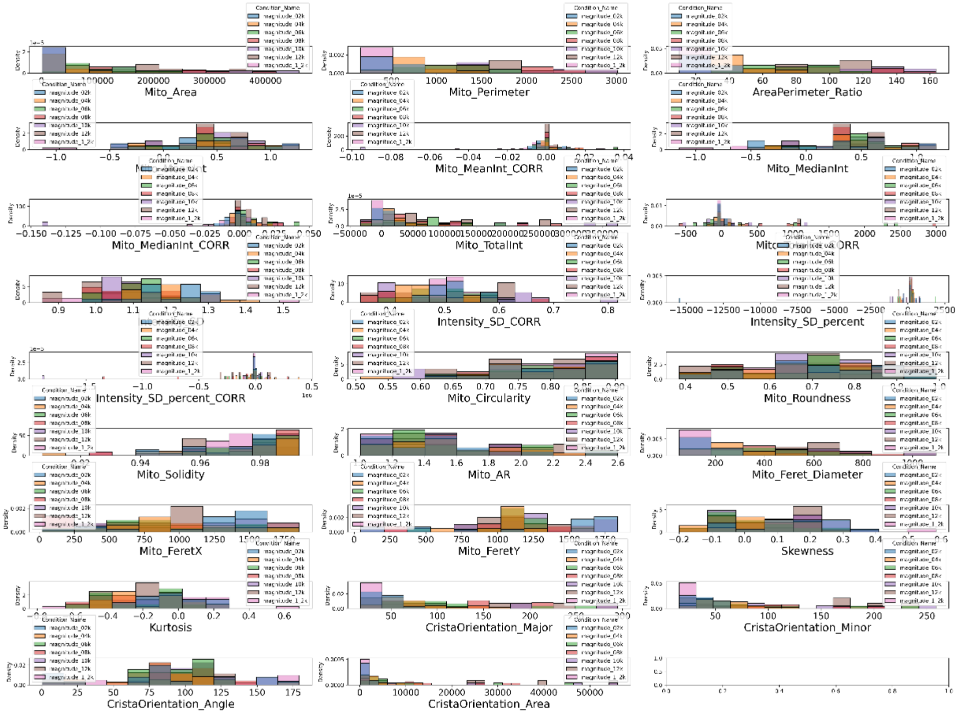
  A density plot is used to represent the distribution of each condition, superimposed on each other. The line represent the kernel density estimator smoothed distribution of density.



Density distribution of mitochondria measurements (one distribution per measurement), according to the magnitude used for EM acquisitions.

- **Histogram distribution**

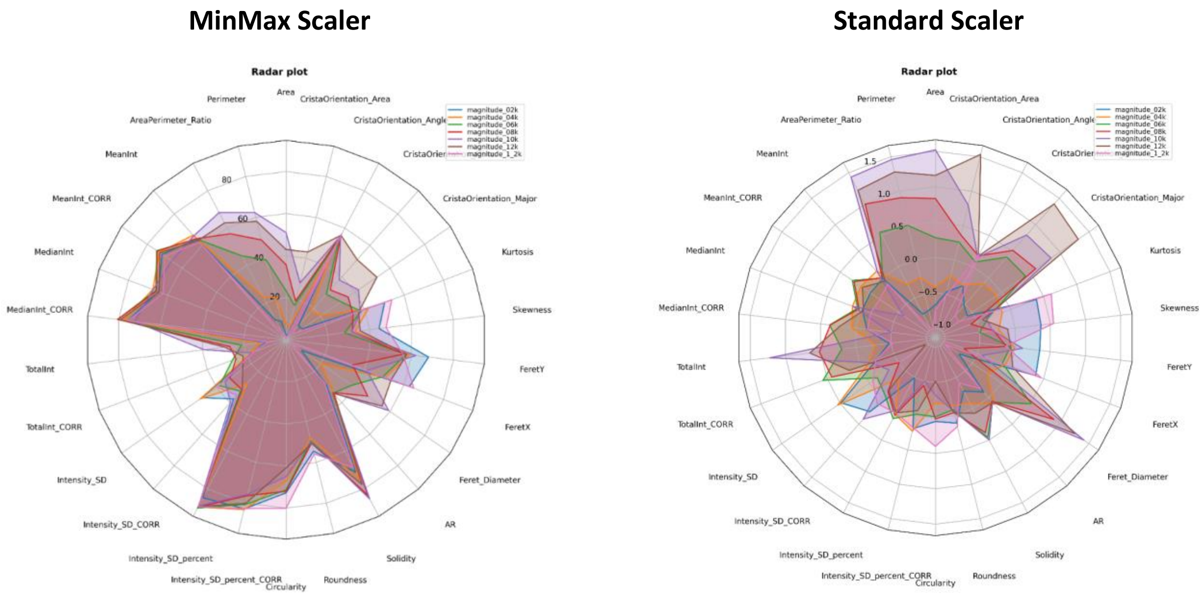
Similar to density plots, histograms are used to represent the density of the distribution of each condition, superimposed to each other. It depicts the distribution with more precision, as it is not smoothed by a kernel density estimator.



Histogram distribution of mitochondria measurements (one distribution per measurement), according to the magnitude used for EM acquisitions.

- **Radar plot distribution**

A radar plot (or radar chart) represent all the features of each condition on a single radial plot for comparison. To compare between features, two scaler have been used: a MinMax Scaler, rescaling all values between 0 and 1 for all features, and a Standard Scaler, substracting the mean value of each feature and dividing by its variance in order to have a pseudo-gaussian distribution centered on 0 with unit variance.

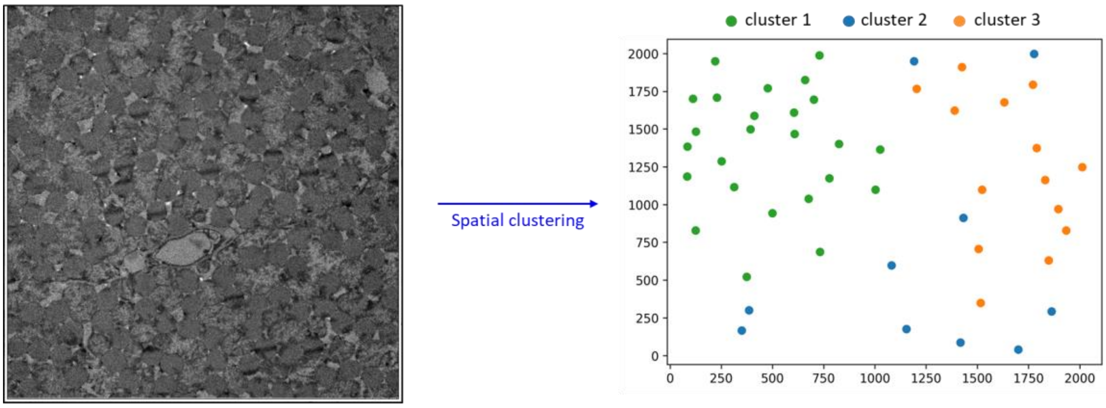


Radar plot distribution of mitochondria measurements according to the magnitude used for EM acquisitions.

- **<u>Spatial clustering distribution</u>**

The spatial clustering plot simply represent the physical position of the centroid of each mitochondria on a two-dimensional plot. A density-based clustering algorithm, HDBSCAN, is used to automatically capture groups of mitochondria, and each of them are averaged for each variable to be compiled in an output csv file.

!! Spatial clustering will increase calculation time !!



Pre-processed TEM image

(X,Y) projection of Mitochondria's centroid

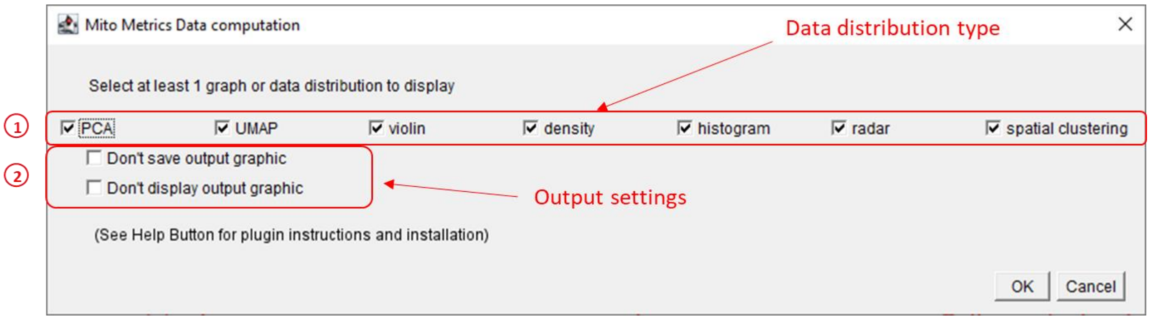
- **Others data computation settings**

  - Don't save output graphic

    Graphs and data distributions are displayed during the analysis, but not saved in the output folder

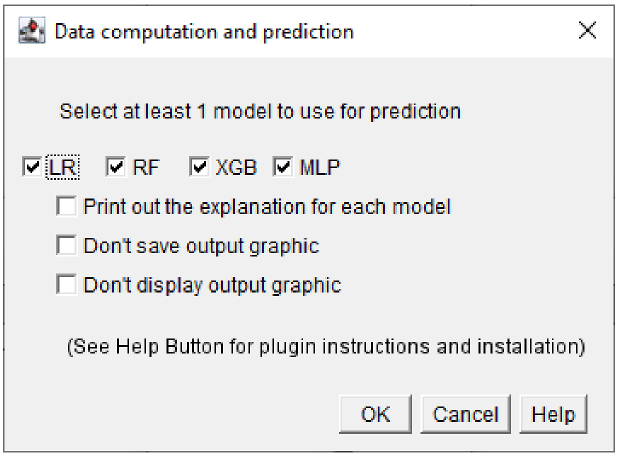
  - Don't display output graphic

    Data graphs and distributions are saved in the output folder, but not displayed during the analysis



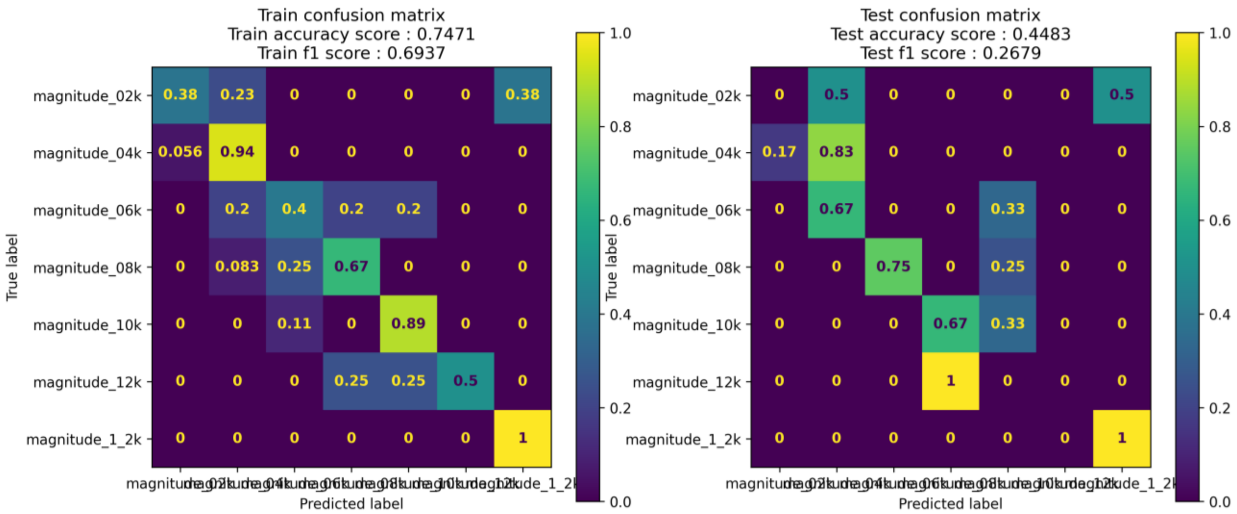
- Setting the analysis: Machine learning models selection

We developed a machine learning (ML) module with predictive analytic tools in order to determine how any given experimental condition would influence mitochondrial morphology and ultrastructure. In the next window, you have to select at least on machine learning model to use for the predictive analysis.



- **<u>Confusion Matrix</u>** (LR, RF, XGB, MLP)

  A confusion matrix presents the predictive performances of a model (see below). The cell (i,j) represents the proportion of mitochondria of true class j that have been predicted of class i by the model. The diagonal (from top-left to bottom-right) represents the correct predictions. Train and test performances have been separated for overfitting assessment, and models performances in term of precision and f1-score are displayed above each matrix.



Four models are proposed for the computing of confusion matrix:

Linear Regression model (LR): tree-based algorithms

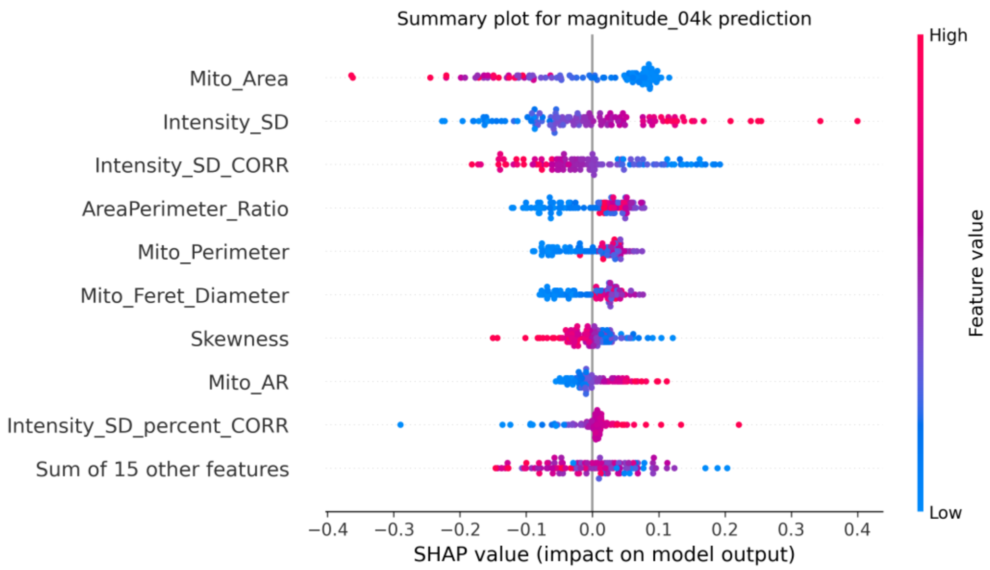
Random Forrest model (RF): tree-based algorithms

XGBoost model (XGB): tree-based algorithms

MultiLayer Perceptron model (MLP): neural network

- **Summary plots (explanation option):**

A summary plot represents the explanation of a model, for a given class (see below). The feature are displayed from the most important (at the top) to the least important (at the bottom). Only the 14 most important features are displayed, and if there are more, they are summed together on a 15th axis. On each axis, each dot represent a mitochondria, its position on the x-axis represent the contribution of the associated feature to the prediction, and its color represent its relative feature value (red represents high values, while blue represents low values). For binary prediction, a single summary plot is displayed, because the summary plot for the second class is the exact opposite of the first one.



SHAP values generated with Linear Regression algorithms (LR), for the magnitude_04k condition

To compute summary plots, activate *print out the explanation function*

- **Others data computation settings**

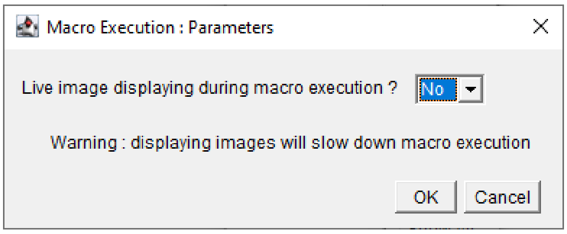
- Don't save output graphic
  Graphs and data distributions are displayed during the analysis, but not saved in the output folder

- Don't display output graphic
  Data graphs and distributions are saved in the output folder, but not displayed during the analysis

3- USER INTERACTION DURING LIVE APPLICATION RUNNING

- Setting live image display

A batch mode is proposed, that allows masking image display during application execution.

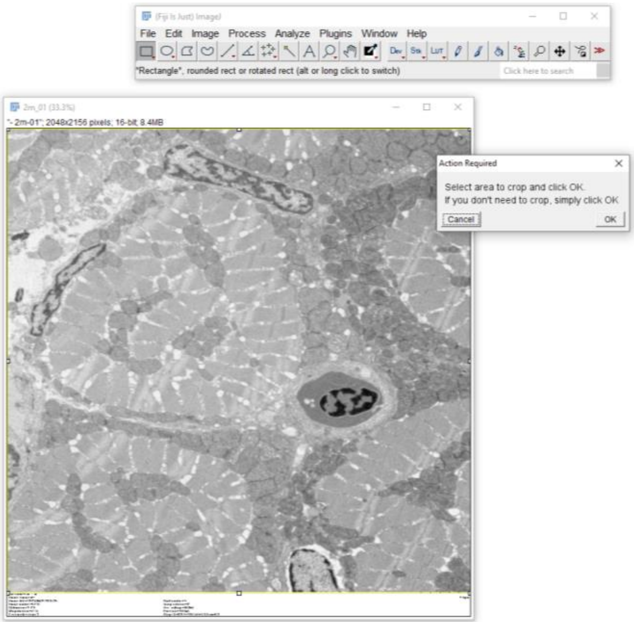


Selecting *live image displaying* (*Yes* settings) may slow down the application execution

- Setting image size reduction of input EM images

Before running mitochondria segmentation and morphological analysis, you can crop your input EM images for each condition separately:

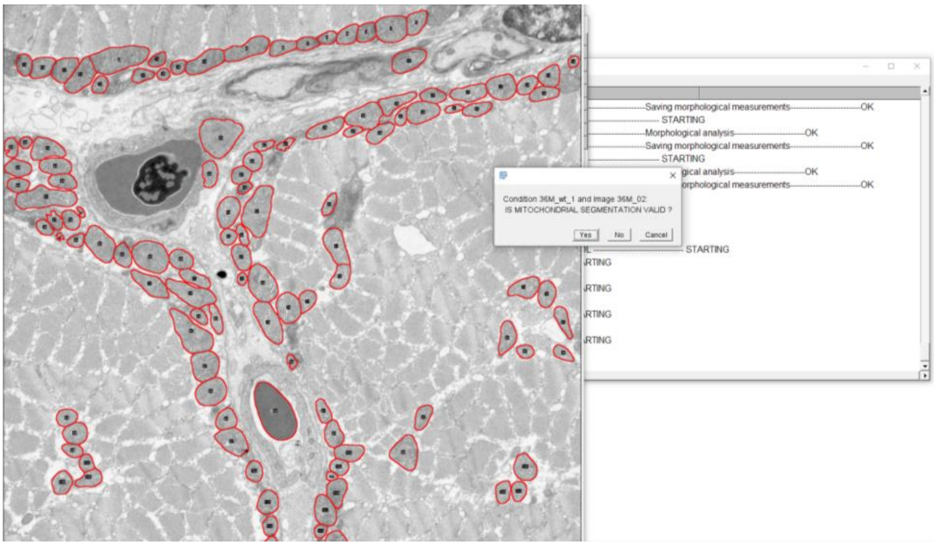
- **If it is necessary, click on *Yes***. Then, **select an area of interest to keep** in your image using the *rectangle tool* from Fiji application. Once the region of interest is selected, click *ok*. The **application will crop all images of the condition, in the same way**.



- **If it is not necessary, click on *No***. The application will continue without cropping images.

- Live quality control of Mitochondria segmentation

Once mitochondria detection is done, validate or invalidate mitochondria segmentation for each input EM image of each condition, as shown below:



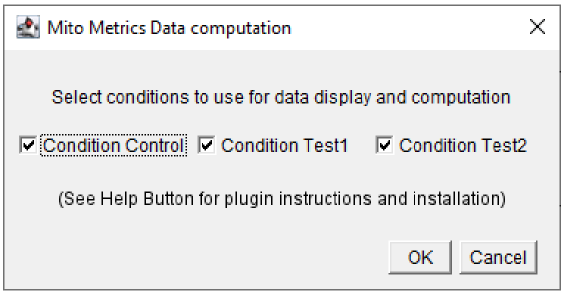
- **<u>Click on *No* button</u>** if you consider that mitochondria **detection is not good**.

- **<u>Click on *Yes* button</u>** if you consider that mitochondria **detection is good**.

Once segmentation validation is done, the application will continue using all images and all mitochondria but the invalid ones

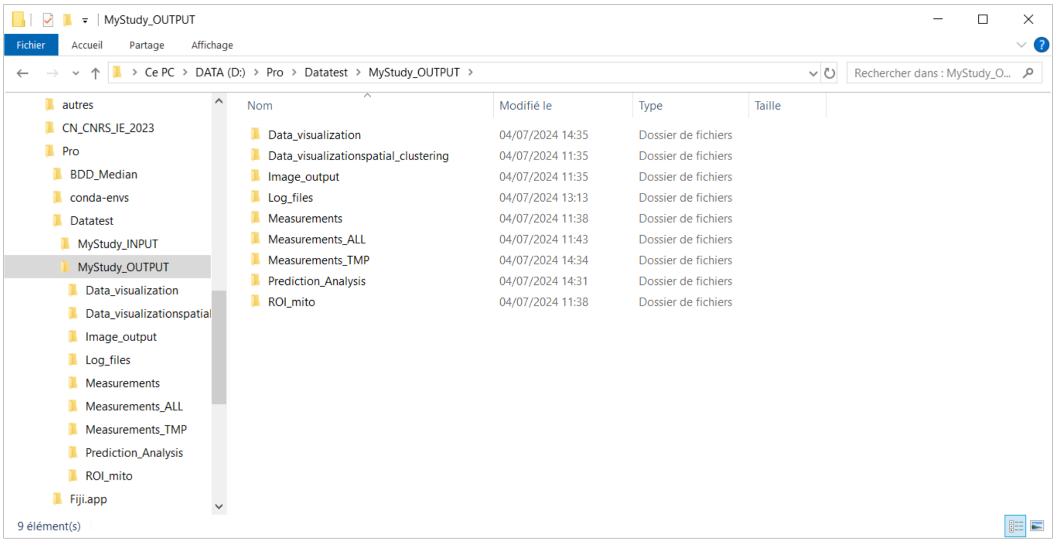
- Selecting conditions to use for data display & prediction analysis

Select conditions to use for data visualization, and those for data prediction. You must select at least one condition for data visualization, and two conditions for data prediction (see below)



EMITO-METRIX OUTPUT DESCRIPTION

In this section, we describe folders, images, measurement files and data distributions saved in the output folders.

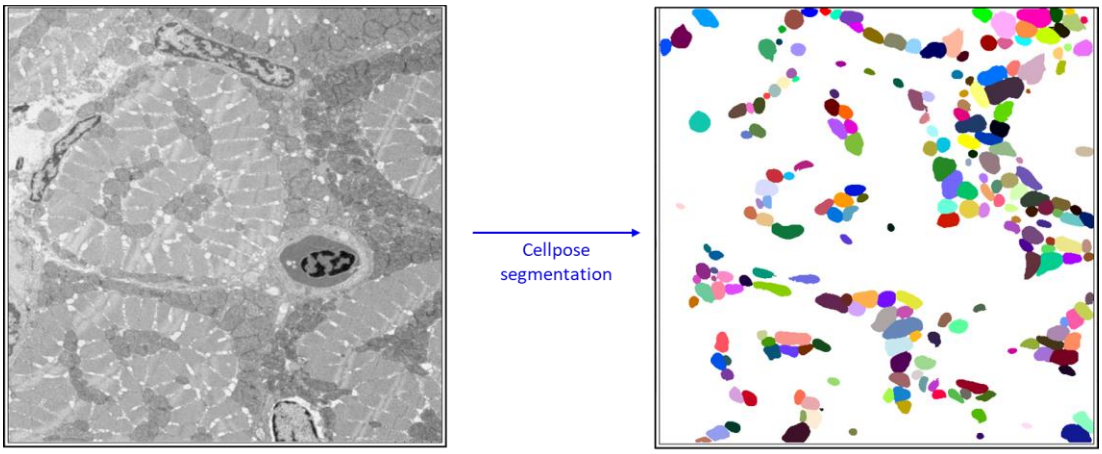


- ***Image_output* folder**

- *processed_images*: normalized & cropped EM images

- *cellpose_mask*: label maps containing segmented mitochondria

One value (i.e. one region of interest ROI) per mitochondria detected

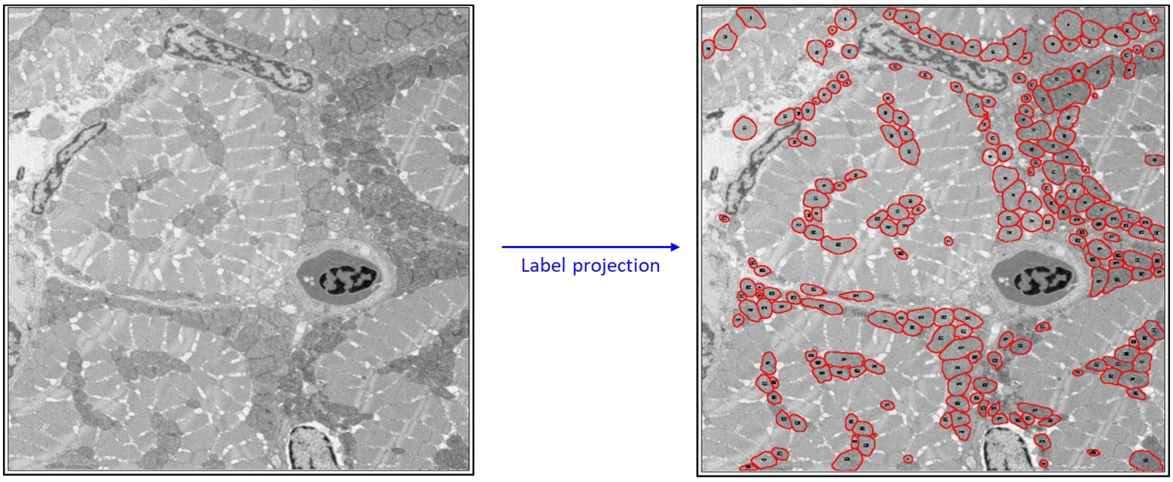


Pre-processed TEM image

Label map of mitochondria

- *control_quality_mask*: projection of mitochondria segmentation and normalized EM image

Used for the quality control of mitochondria detection



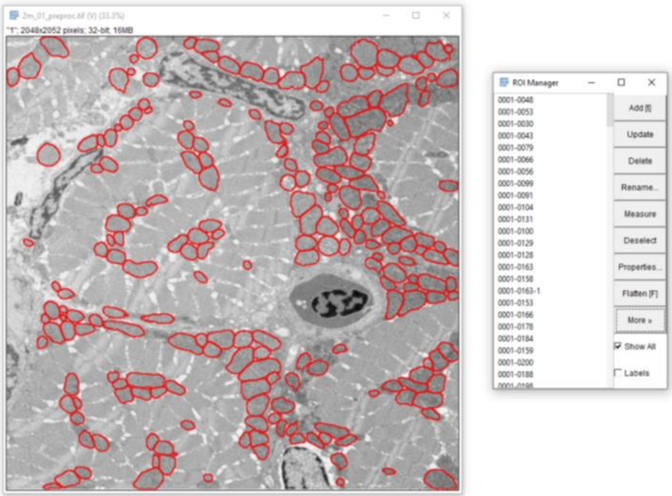
Pre-processed TEM image

Mitochondria label projection

- ***ROI_mito* folder**

Label regions of mitochondria detected with Cellpose or custom models. These regions of interest (ROI) files are opened and displayed using the *Fiji-ROI manager tool*

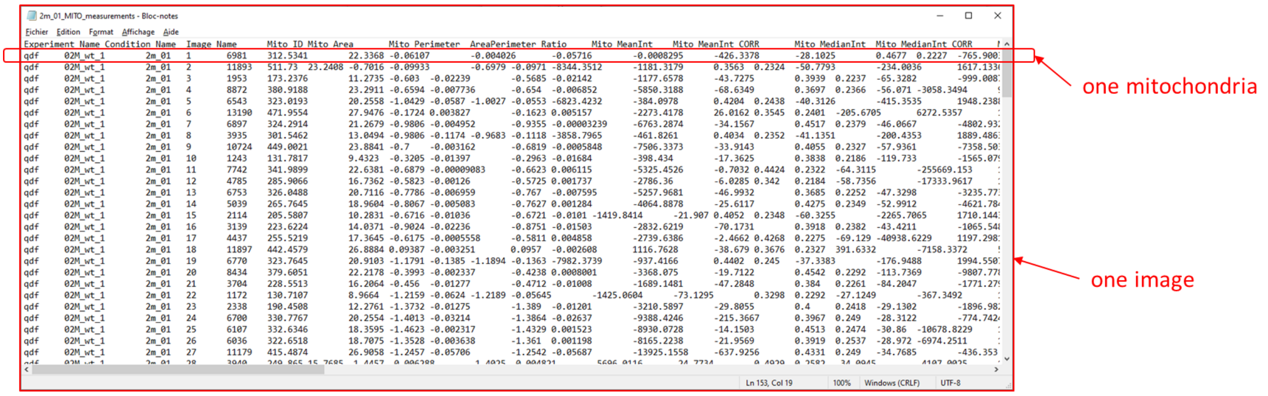
Drag and drop your EM image in Fiji, then open the ROI file using Analyze/Tools/ROI manager



- ***Measurements* folder**

Morphometrics files containing morphological and texture measurements (see [here]() for a detailed description of mitochondrial measurements).

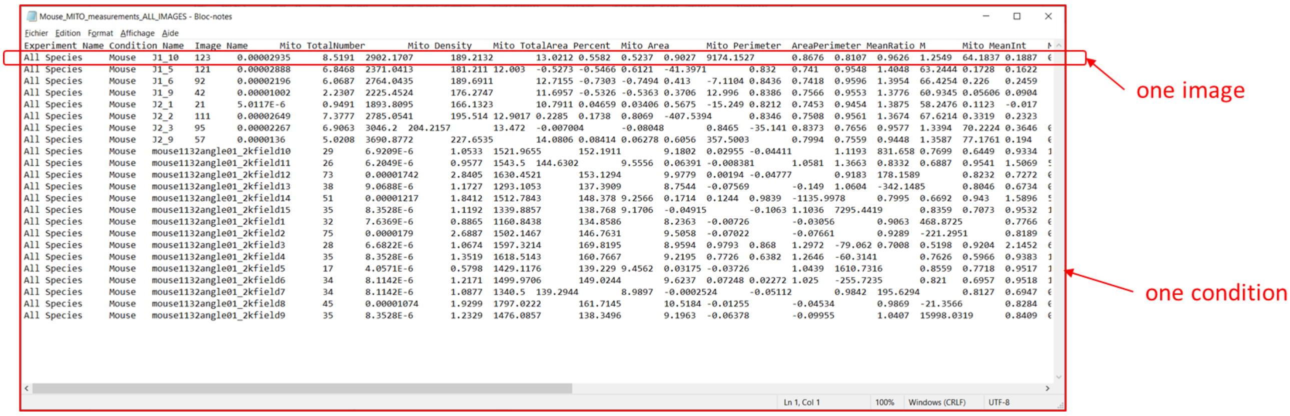
One text file per image and one line per mitochondria detected



- ***Measurements_ALL* folder**

Contains average morphological and texture measurements for each image (see [here]() for a detailed description of mitochondrial measurements)

One text file per condition, and one line per image



- ***Measurements_TMP* folder**

Temporary folder used during application execution

- ***Data_visualization* folder**

Folder containing graphs and distributions for data visualization

__density.png__: density distribution of mitometrics for each condition

__histogram.png__: histogram distribution of mitometrics for each condition

__MinMaxScaler_radar_plot.png__: radar plot of MinMax rescaled mitometrics for each condition

__StandardScaler_radar_plot.png__: radar plot of Standard rescaled mitometrics for each condition

__PCA_Condition_Name.png__: PCA distribution of mitochondria according to the conditions

__UMAP_Condition_Name.png__: UMAP projection of mitochondria according to the conditions

__violin.png__: violin distribution of mitometrics for each condition

- **Data_visualizationspatial_clustering folder**

Folder containing spatial clustering of mitochondria detected for each image (*spatial_clustering* file), as well as density graph (*density* file) and mean morphometrics (*clusters* file) computed for each cluster

- ***Prediction_analysis* folder**

Folder containing SHAP values (*beeswarm* file) and confusion matrix (*confusion_matrix* file) for each machine learning algorithms used.

- **Log_files folder**

Folder containing application settings *(Users_general_parameters.txt)* and segmentation settings *(Users_Mitochondria_size.txt)*.