

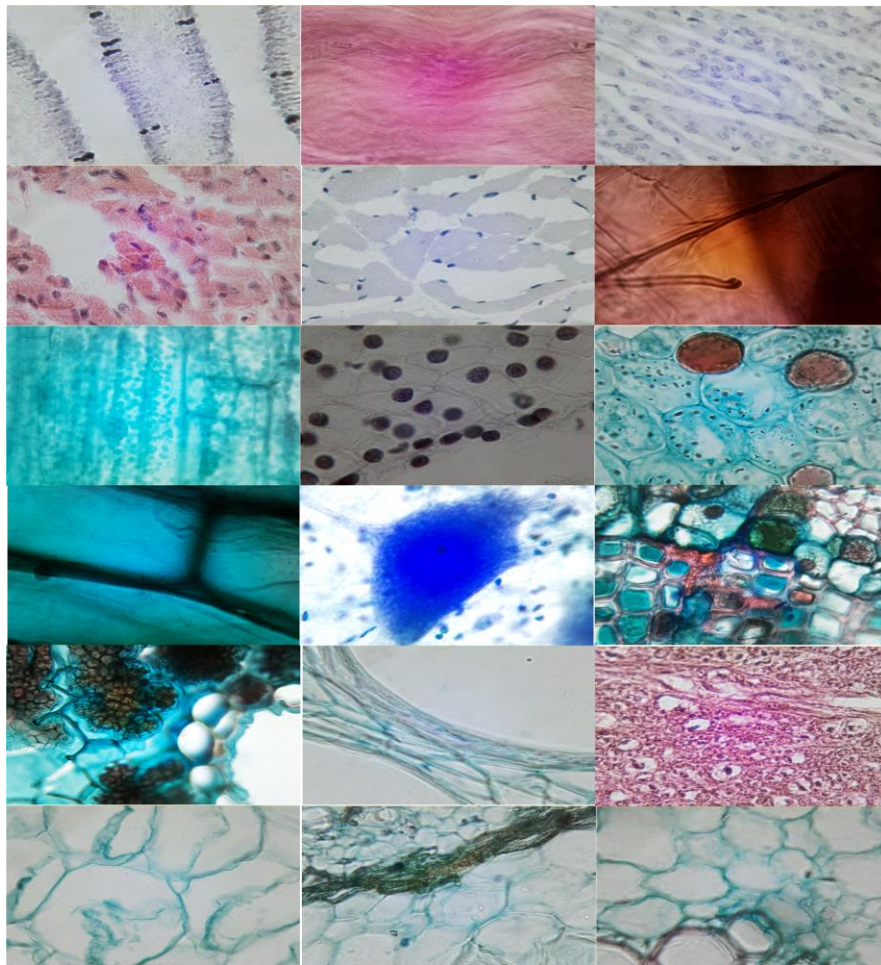
## OPEN-BIOset's

### Contact Information:

For all details and assistance regarding the datasets and the project, please reach out to Dr Akhil Kallepalli<sup>1</sup> through [Email](#) and/or [LinkedIn](#).

### Contents:

The OPEN-BIOset's contain microscopic images of 19 different biological samples to aid the advancement of open-source, machine learning-based microscopy. Containing both animal and plant tissues - such as pumpkin stem, pig motor nerve, and dog skeletal muscle – the OPEN-BIOset's enable the development of novel techniques for microscope image classification. OPEN-BIOset was established using a x40 magnification and OPEN-BIOset 2.0 utilises a x100 magnification. A selection of images from OPEN-BIOset 2.0 is presented below in Figure 1.



---

<sup>1</sup> School of Physics and Astronomy, University of Glasgow (Kelvin Building, Glasgow G12 8QQ, UK)

Figure 1: An assortment of images from 18 of the categories.

The primary motivation of these dataset's is not only to illustrate the potential of the OpenFlexure microscope and the advantages of open-source science but also to provide a dataset for future machine learning algorithm testing as is the case with datasets released in other fields such as autonomous driving and medical imaging.

### Layout:

The 19 biological samples were imaged using a 3D printed OpenFlexure Delta Stage microscope, saved to a JPEG format, and have been grouped by category into directories. Each directory contains *c.100 examples* and an additional directory, titled "Backgrounds", contains a corresponding background image for each biological sample (only included for OPEN-BIOset). A full list of sample categories and corresponding exposure settings are presented in Table 1.

Background images for "Dense Connective Tissue", "House Bee Mouth Parts", "Pig Motor Nerve" and "Rabbit Spinal Cord" are currently not included in the dataset but can be replicated by selecting another background image with the appropriate exposure value.

Table 1: Contents of dataset and exposure settings for OPEN-BIOset.

Categories	Number of images	Exposure (J/m <sup>2</sup> )
Backgrounds	15	
Coprinus Mushroom Set	100	476
Dense connective tissue	100	1080
Dog cardiac muscle	100	1080
Dog oesophagus	100	1080
Dog skeletal muscle	100	1080
House bee mouthparts	100	1474
Hydrilla verticillata leaf	102	1474
Lillium ovary	100	1040
Nymphaea of aqstio stem	100	1080
Onion epidermis	100	24969
Pig motor nerve	100	1080
Pine leaf	100	1474
Pine stem	100	4980
Pumpkin stem	100	1080
Rabbit spinal cord	100	1080
Sunflower stem	100	1040
Tilia stem	100	1080
Young root of broad bean	100	1040
Zea stem	100	1040

## Optics:

The optics module for the microscope that created OPEN-BIOset consisted of a Raspberry Pi camera V1.3, capable of 5MP resolution (the original Pi camera lens has been removed) and an x40 RMS threaded microscope objective lens. A white LED, a 13 mm x 7.5 mm achromatic lens and a diffuser plate make up the illumination module. All resultant images were of size 2560 × 1440 pixels.

The optics module for the microscope that created OPEN-BIOset 2.0 consisted of a Raspberry Pi camera V2.0, capable of 8MP resolution (the original Pi camera lens has been removed) and an x100 RMS threaded microscope objective lens. A white LED, a 13 mm x 7.5 mm achromatic lens and a diffuser plate make up the illumination module. All resultant images were of size 3264 × 2448 pixels.

## Using the OPEN-BIOset's:

Found in the GitHub repository — <https://github.com/rob-arch/OPEN-BIOset.git> — is the Python code containing the classification algorithm.

Once an OPEN-BIOset is loaded into the directory, ready for pre-processing, the code resizes all images to 256 x 256 and ignores the “Backgrounds” folder from further processing.

The classification algorithm, written in Python with TensorFlow with Keras, forms two convolutional layers to identify characteristics unique to each biological sample with max-pooling layers succeeding and a SoftMax output layer to classify the images.

A confusion matrix for each dataset is produced directly from the Convolutional Neural Network. The confusion matrix consists of a row and column for each sample, with ‘Actual Values’ along the x-axis and ‘Predicted Values’ along the y-axis. This results in a diagonal line arising in the middle of the model if the model correctly categorises each sample. Values above the diagonal line represent a False Positive categorisation and values below represent a False Negative. This is an excellent visualisation of the OPEN-BIOset's and the classification algorithms capabilities.

## License:

Images included in this repository are licensed under CC BY-NC-SA 4.0, a full copy of this license be found [here](#).

## Credit:

If these dataset's help your research, please give recognition by citing the dataset directly with its DOI or the accompanying paper.

Contributors: Robert Archibald, Samuel T Westlake, Graham Gibson, Akhil Kallepalli

Archibald R, Gibson G, Westlake S T, Kallepalli A, “*Open-source microscopic solution for classification of biological samples.*” Additional details to be confirmed/updated.