HistoClean Documentation
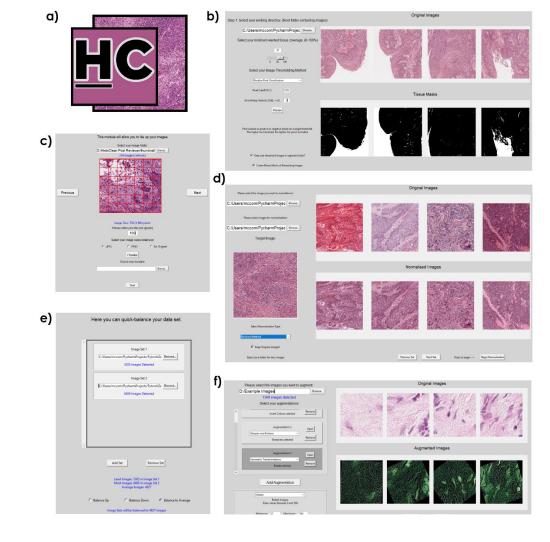
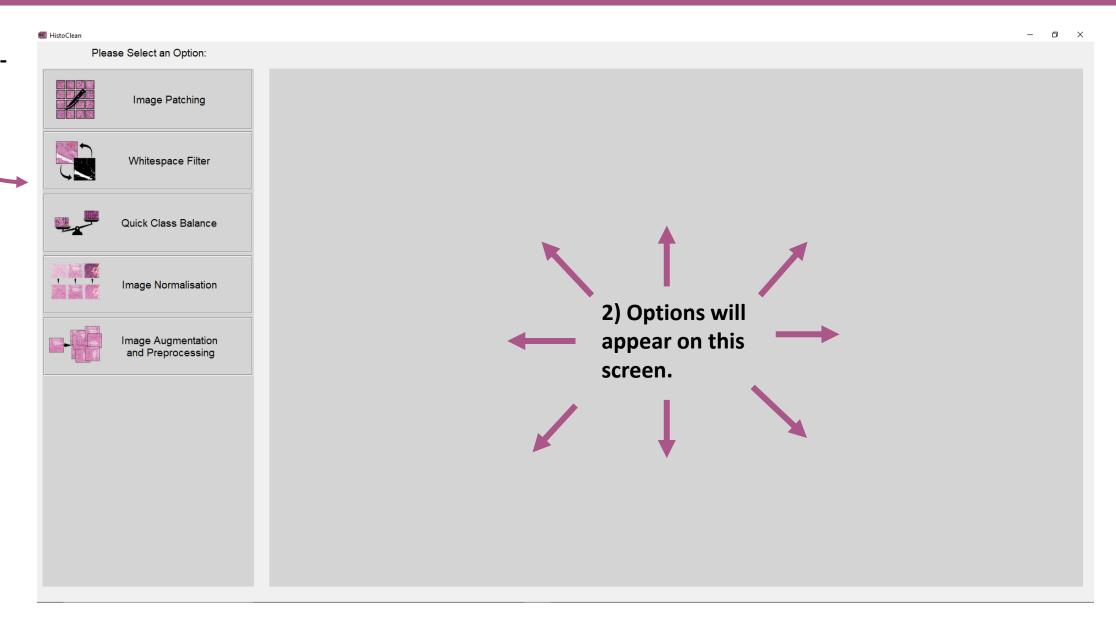Author: Kris McCombe

HistoClean Background

- HistoClean is an image pre-processing application for use in deep learning projects in digital pathology.
- The standalone Graphical User Interface allows users to clean up their image datasets, which gives the best opportunity for developing a robust deep learning model.
- HistoClean **(a)** is split into 5 modules:
    b) Whitespace Removal
    c) Image Patching
    d) Image Normalisation
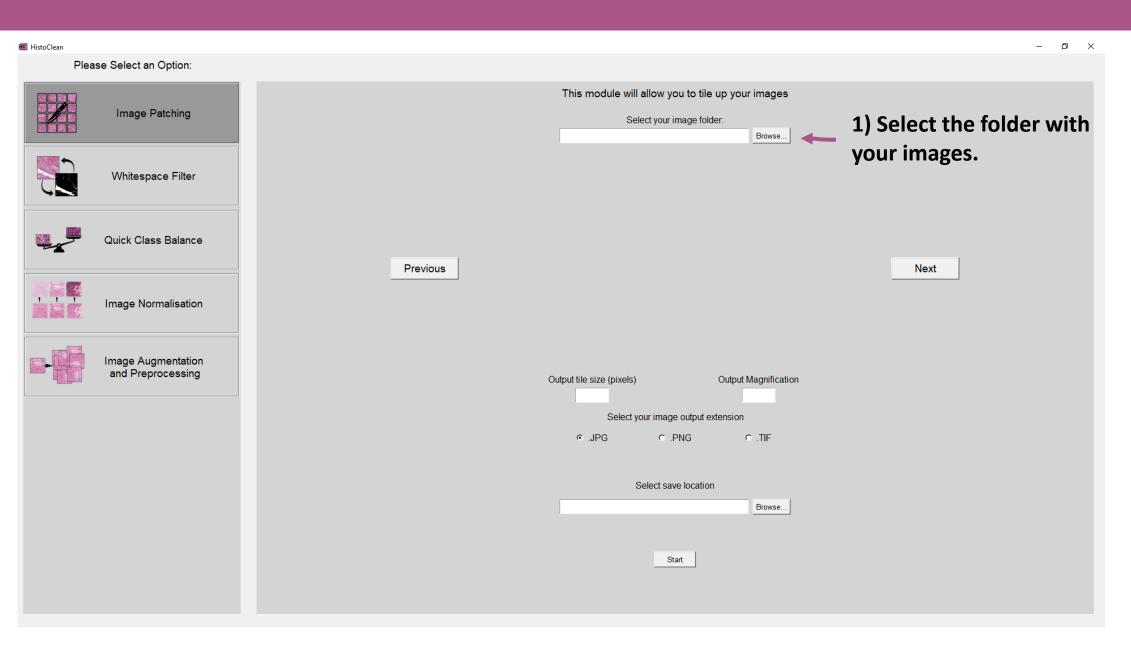   **e)** Quick Balancing
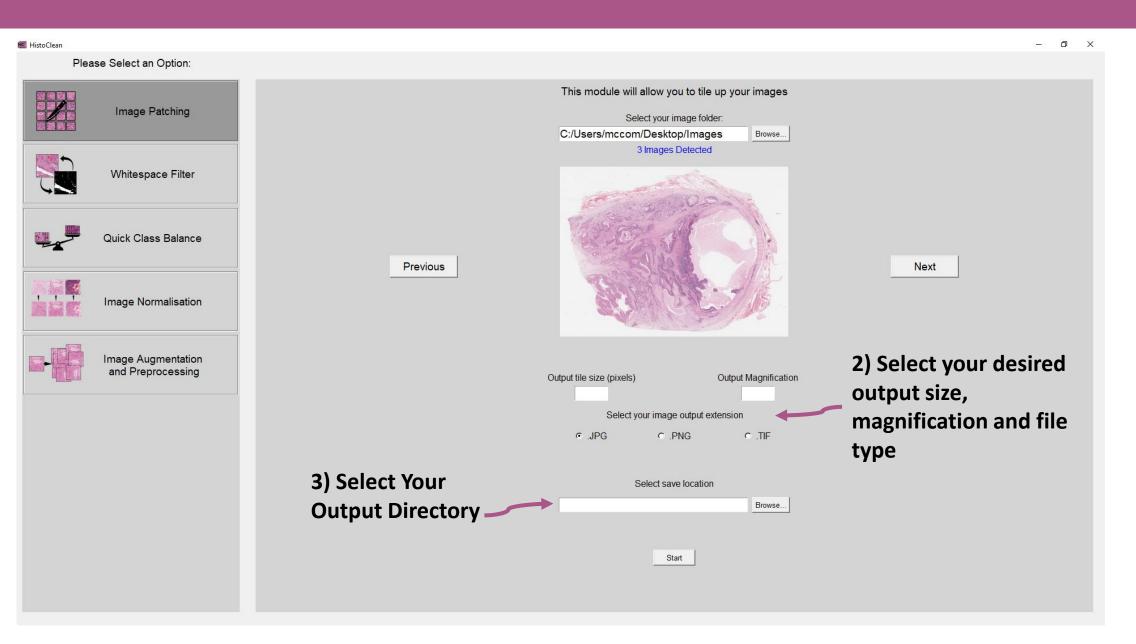   **f)** Image Pre-processing & Augmentation

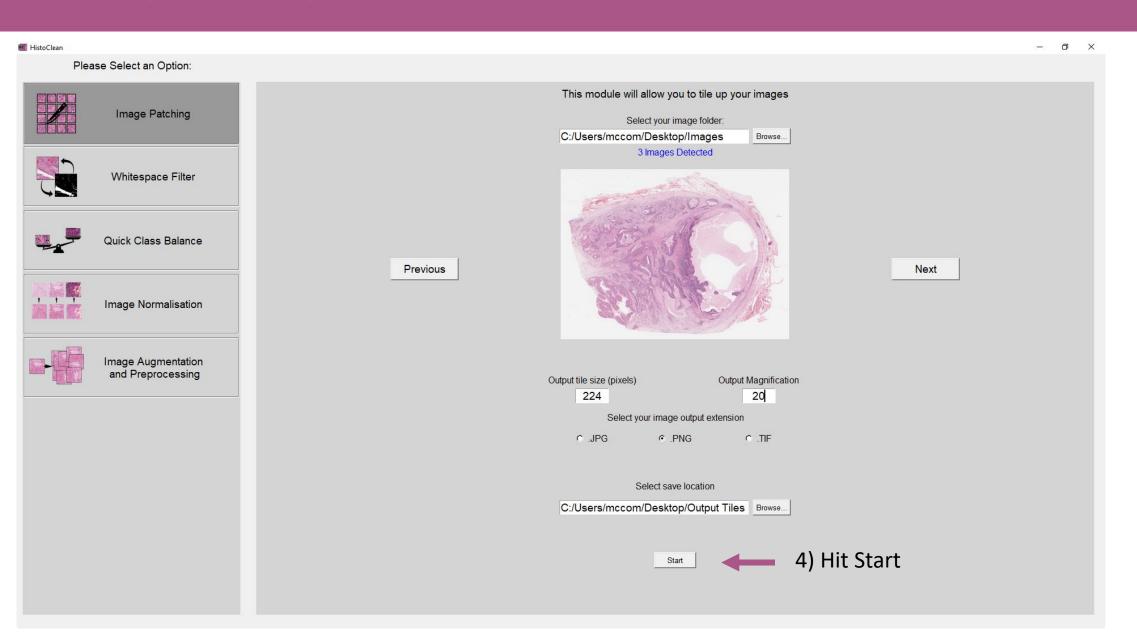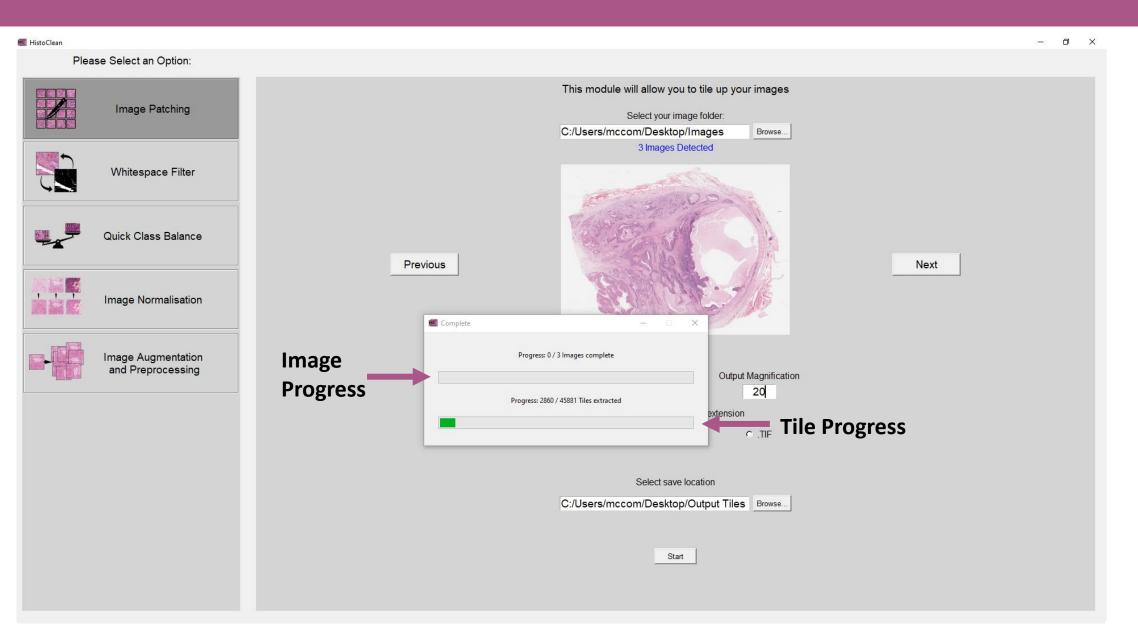# The Interface

1) Select a pre-processing step to apply.

2) Options will appear on this screen.

1) Image Patching

1) Image Patching

1) Image Patching

1) Image Patching

1) Image Patching
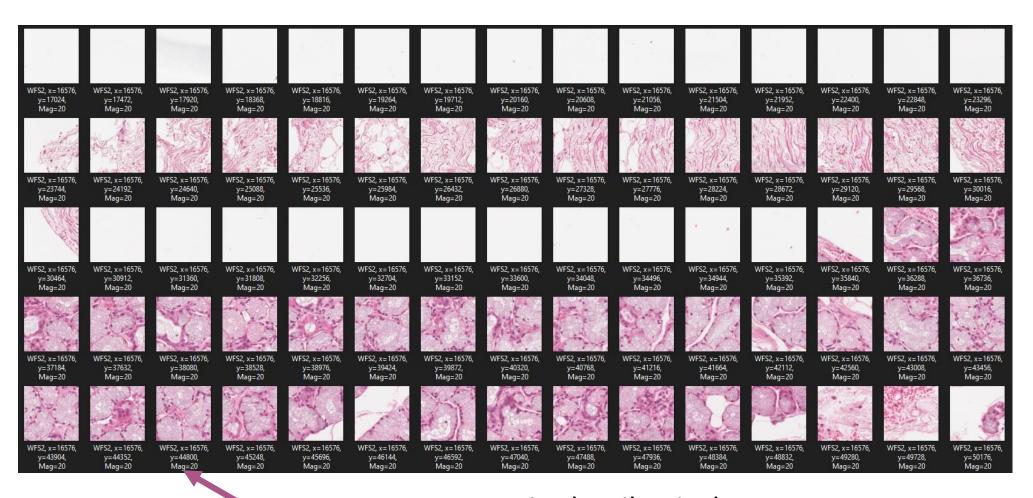
Image names contain the tile pixel coordinates and output magnification.
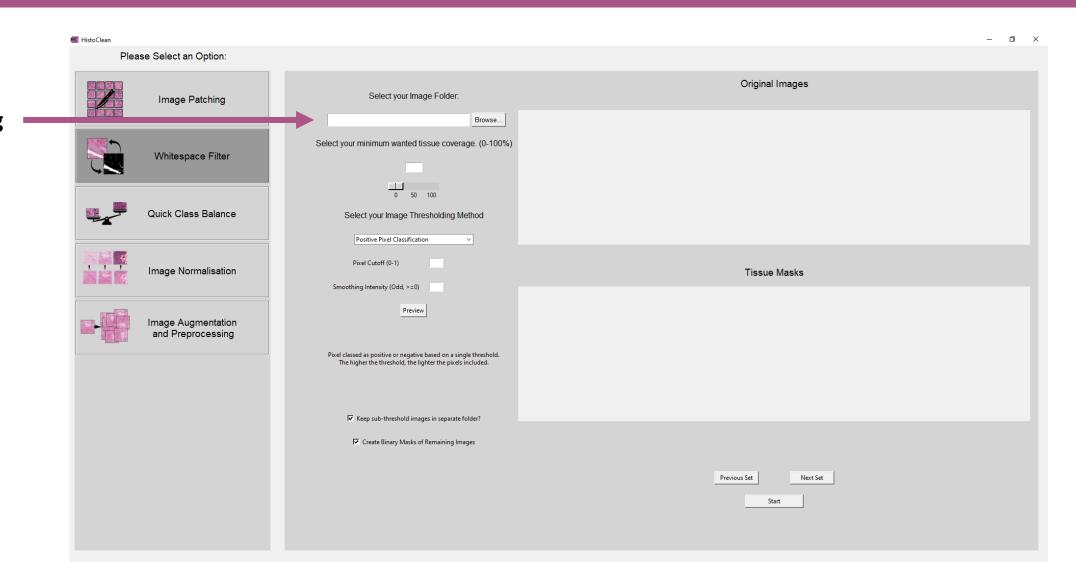
2) Whitespace Filter

The Whitespace Filter option allows you to remove any tiles below a chosen tissue percentage
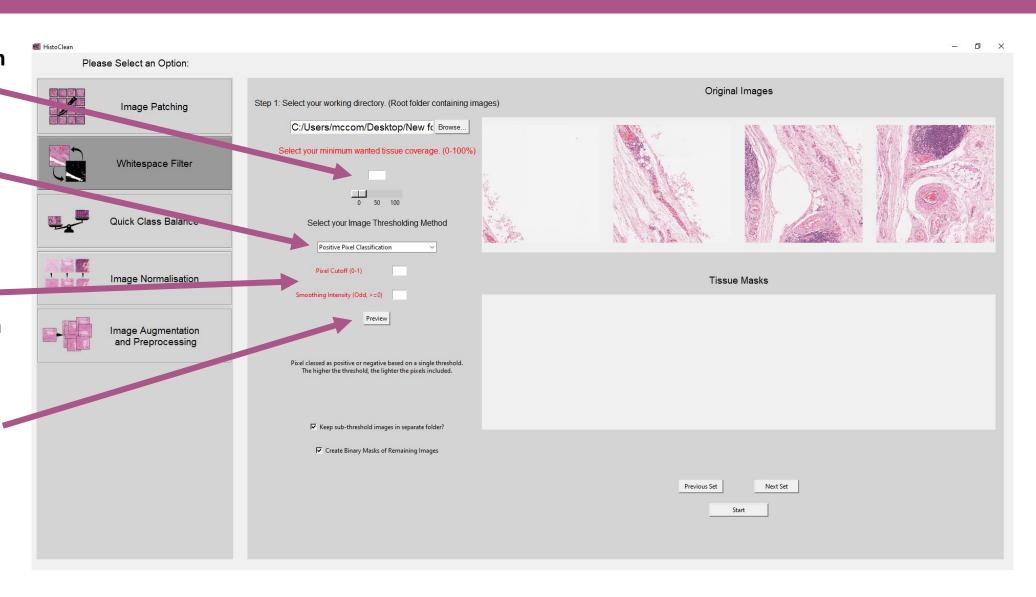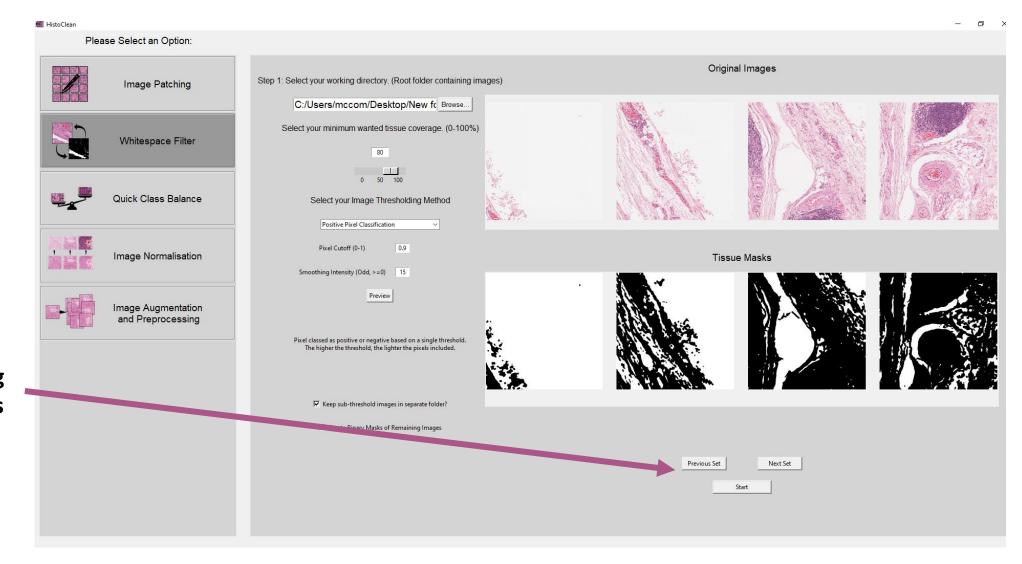
1) Select the folder containing your images

2) Whitespace Filter

2) Select the minimum tissue you want your tiles to have

3) Select the method of pixel classification

4) Select the Positive pixel threshold and smoothing intensity. Smoothing helps fill in small holes in the tissue
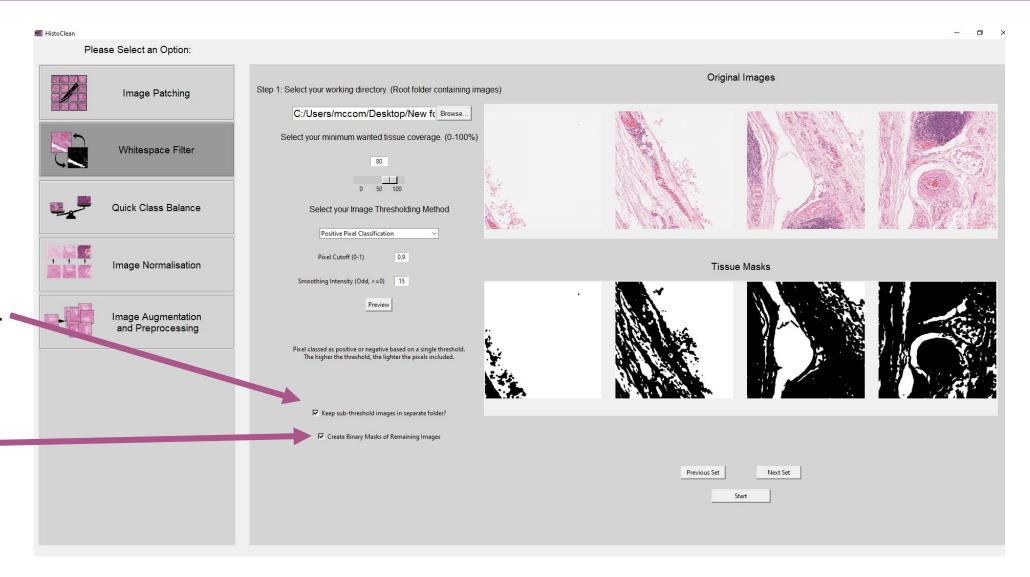
5) Preview the results

6) You can preview different images using the Next and Previous Buttons
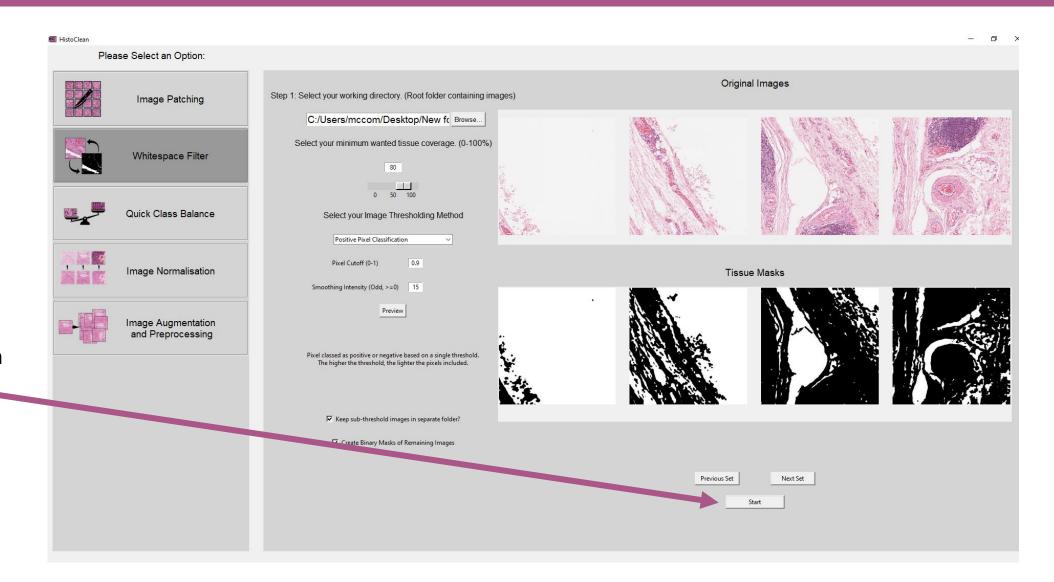
2) Whitespace Filter

7) You can decide whether to keep the sub-threshold images in a separate folder or delete.

8) You can also have binary masks created of the remaining images

2) Whitespace Filter
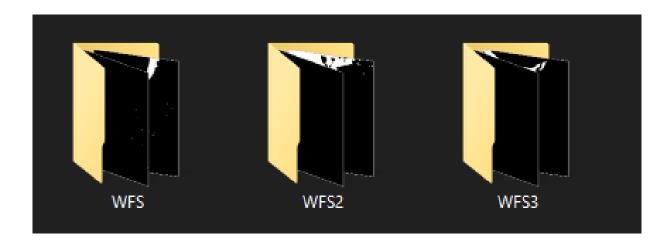

9) Hit Start to begin

2) Whitespace Filter

Two new folders appear, *Masks* and *Removed Images*

2) Whitespace Filter
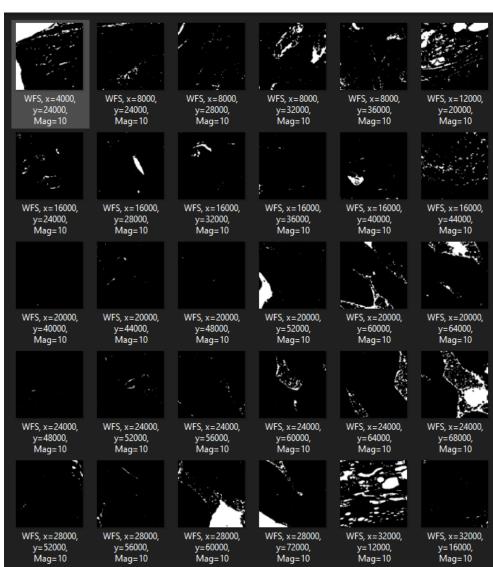
Masks Contains the Tissue Masks for each of the tiles arranged the same as the input folder

2) Whitespace Filter

Masks

Originals
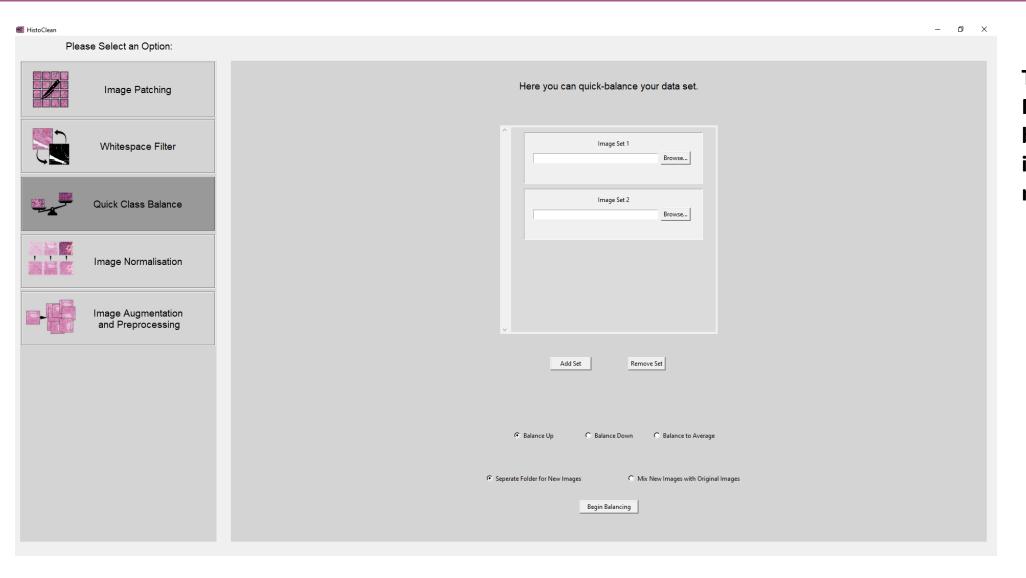
2) Whitespace Filter

The Removed Images Folder contains all the tiles that were below the tissue percentage cut-off
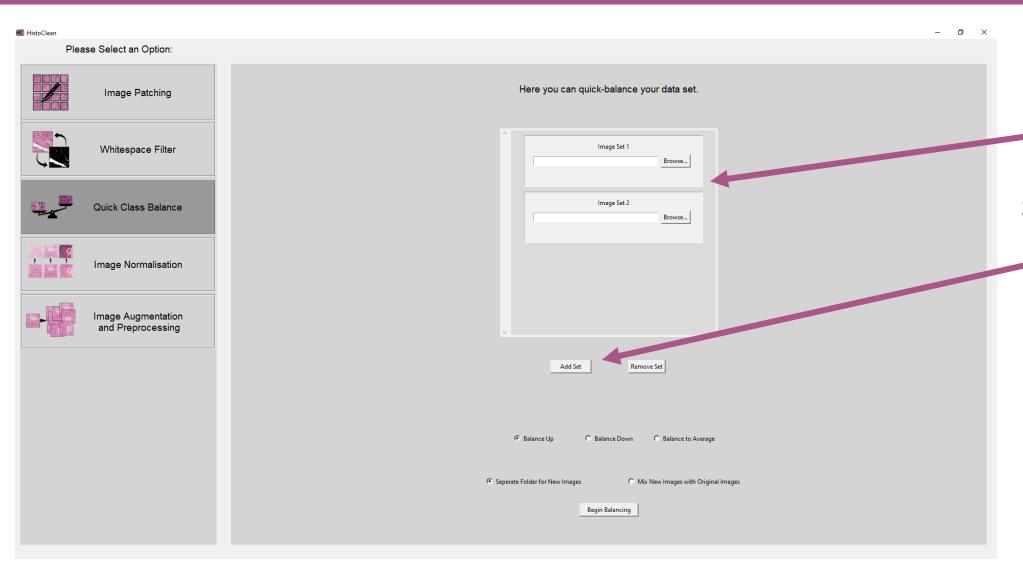
3) Quick Balance

The Quick Balance Module allows you to balance classes in an image set using simple mirroring and rotations.

3) Quick Balance

1) **Add the folder paths to each class in your image set.**

2) **You can add additional classes by clicking the "Add Set" Button**
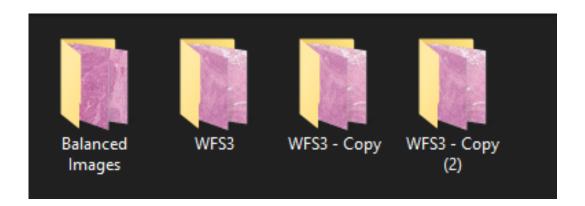
3) Quick Balance

3) Decide to balance to the lowest, highest or average number of images in a class

4) Decide to mix the new images with the old ones, or keep them in a separate folder

5) Click "Begin Balancing"

3) Quick Balance

If you Selected "Separate folder for new images" There will be a new folder called "Balanced Images"
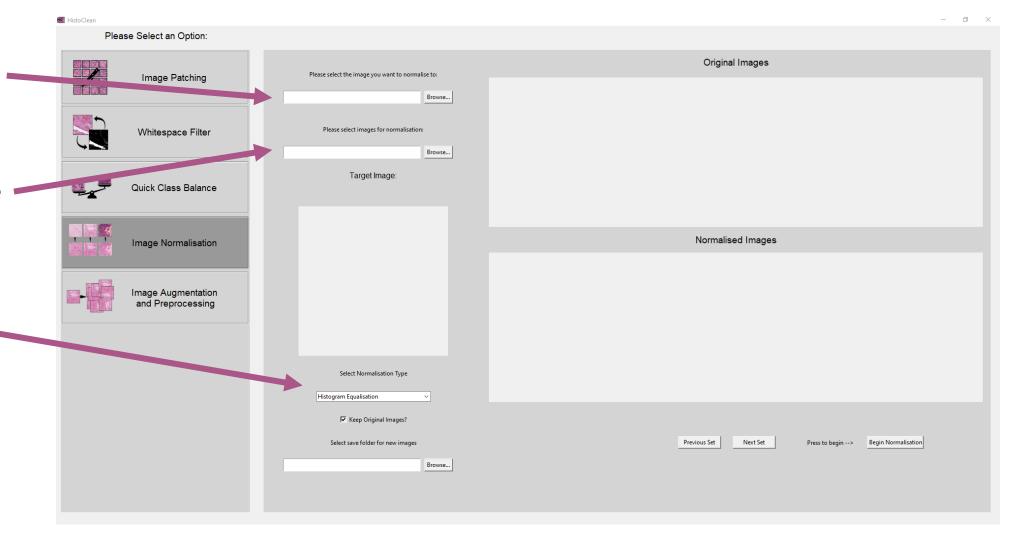
4) Image Normalisation

- **Image Normalisation can be used to make stain profile of the images look this same.**

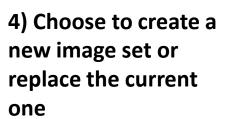- **This helps remove colour bias from the model to be trained.**

4) Image Normalisation

1) Select the image to normalise to

2) Select the folder with the images to be normalised
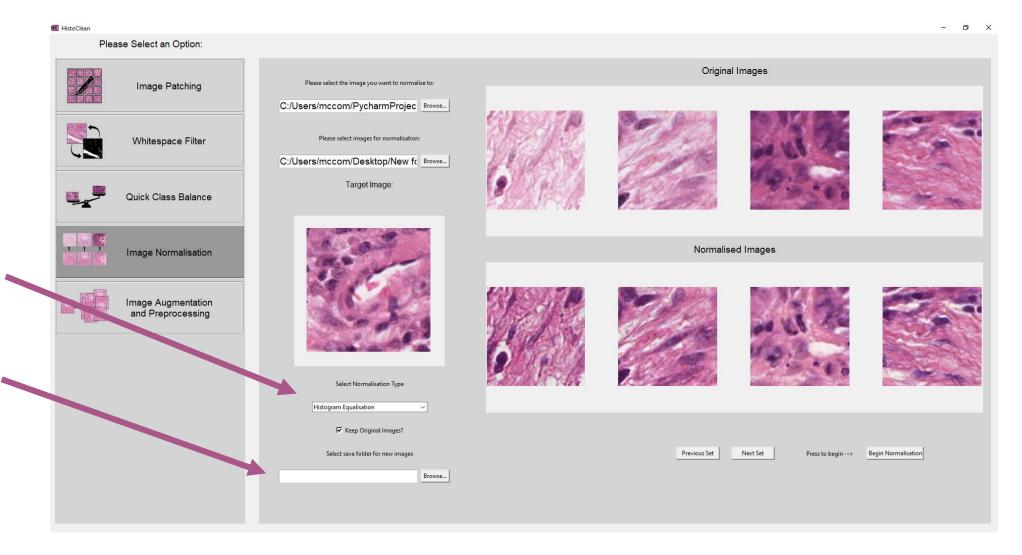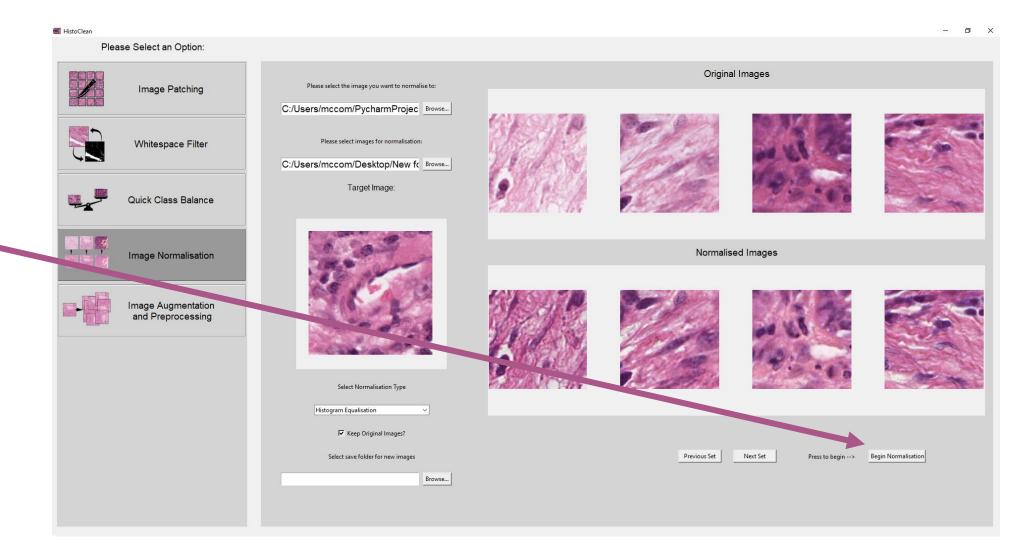
3) Select the normalisation type

4) Image Normalisation

4) Choose to create a new image set or replace the current one

5) Select an output directory if required

4) Image Normalisation
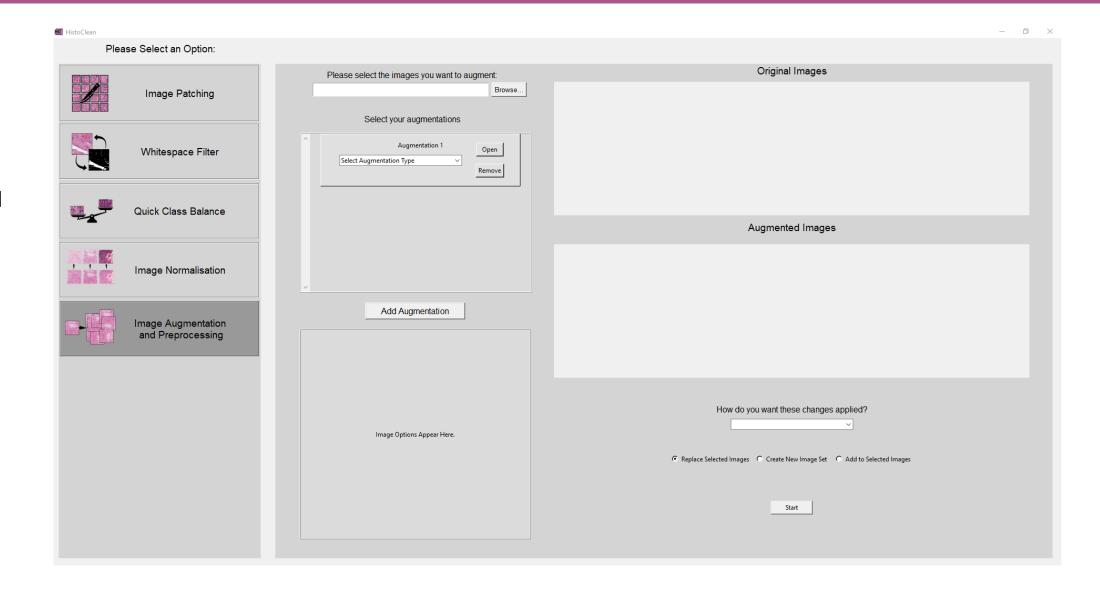
6) Click "Begin Normalisation"

5) Image Augmentation and Pre-processing

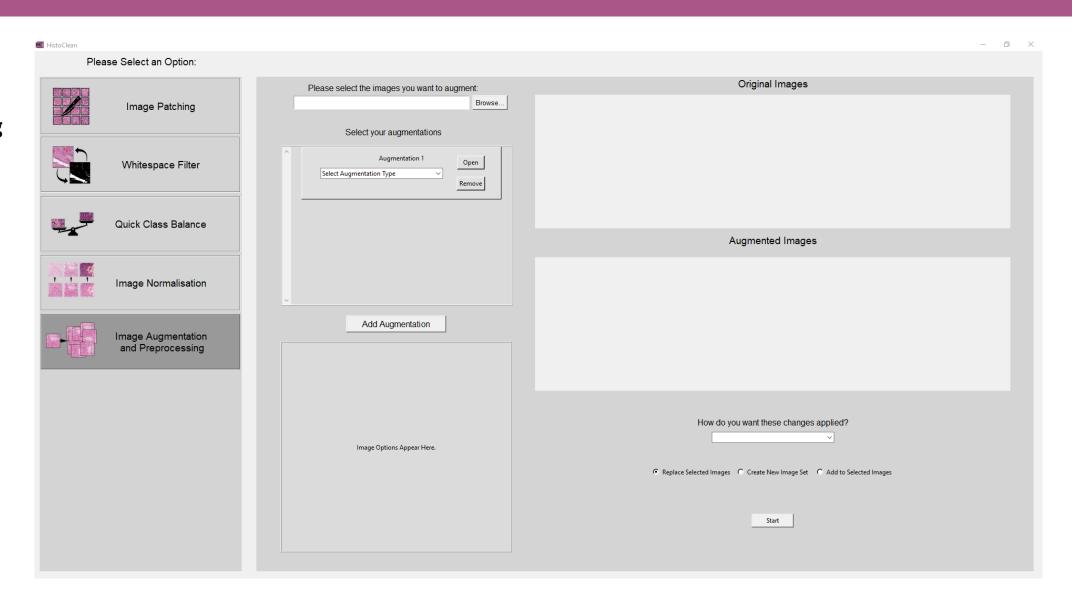- **This module allow you to apply different effects to your images, as well as artificially expand the image set if desired.**

5) Image Augmentation and Pre-processing

1) Select the folder containing the images to augment.
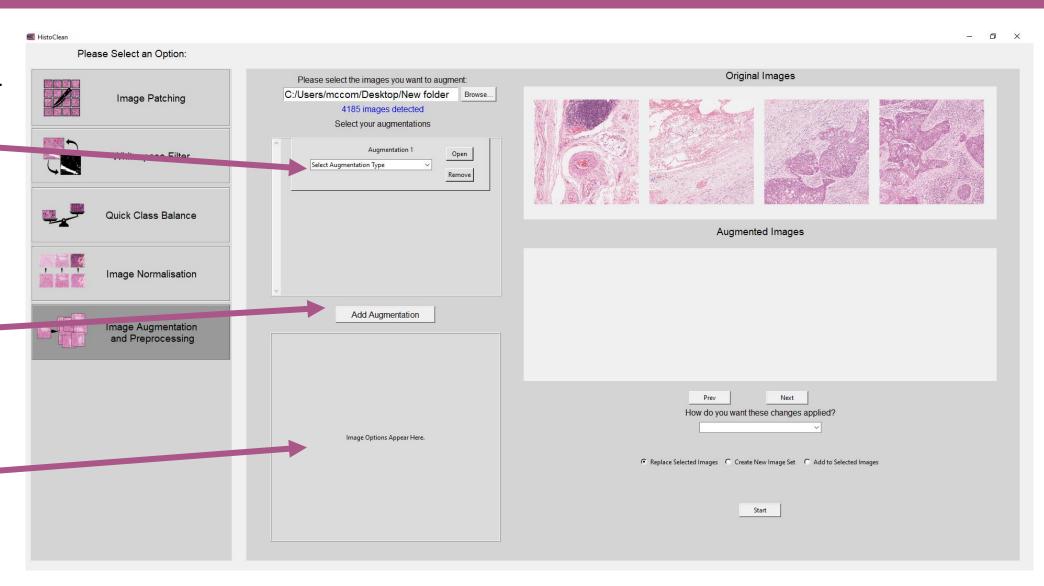
5) Image Augmentation and Pre-processing

2) Select the pre-processing effects to be applied to the images.

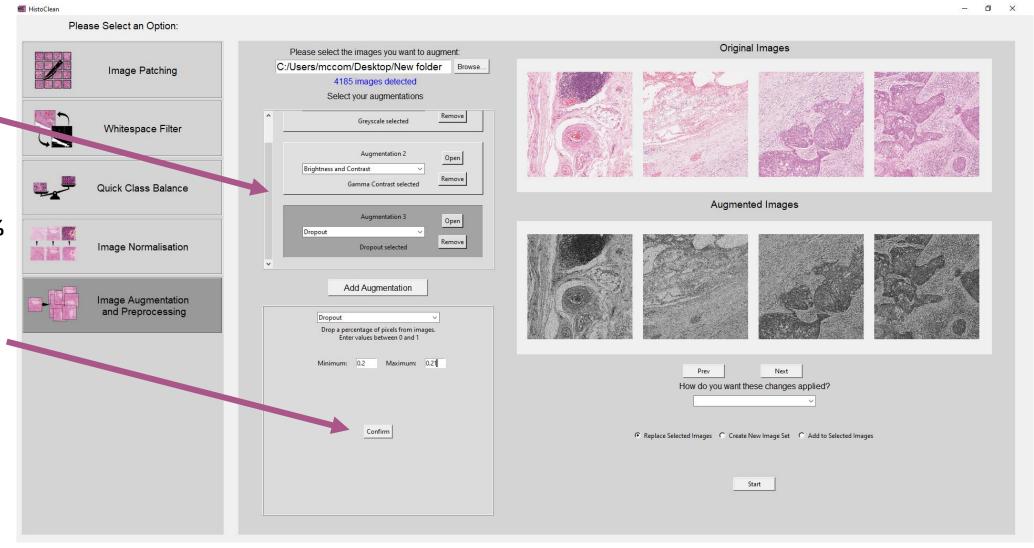**You can add multiple effects by clicking the "Add Augmentation" button**

3) Options Appear here

5) Image Augmentation and Pre-processing

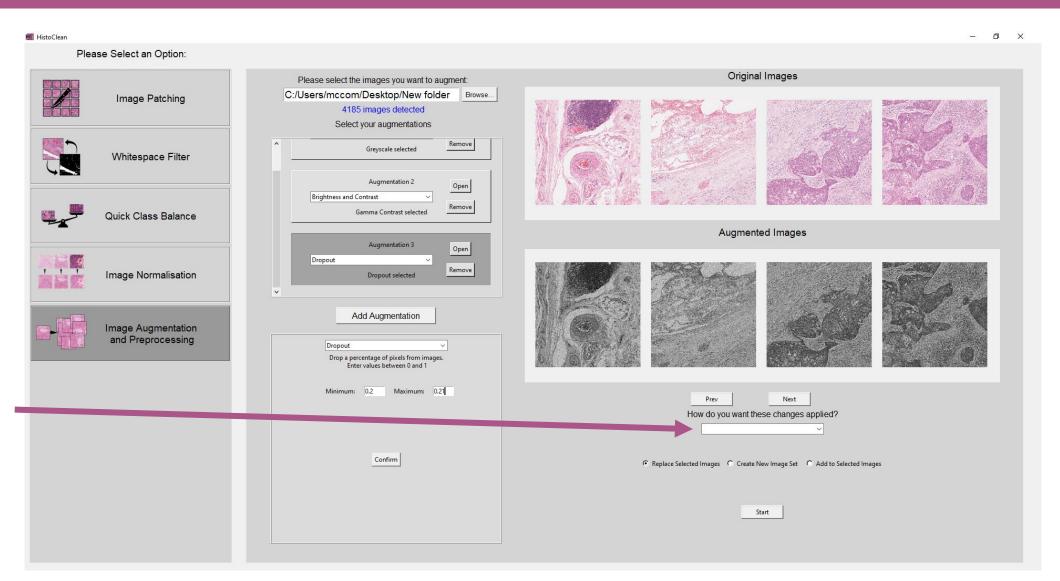**4) In this example, we greyscaled our images, increased the contract and dropped out 20% of the pixels.**

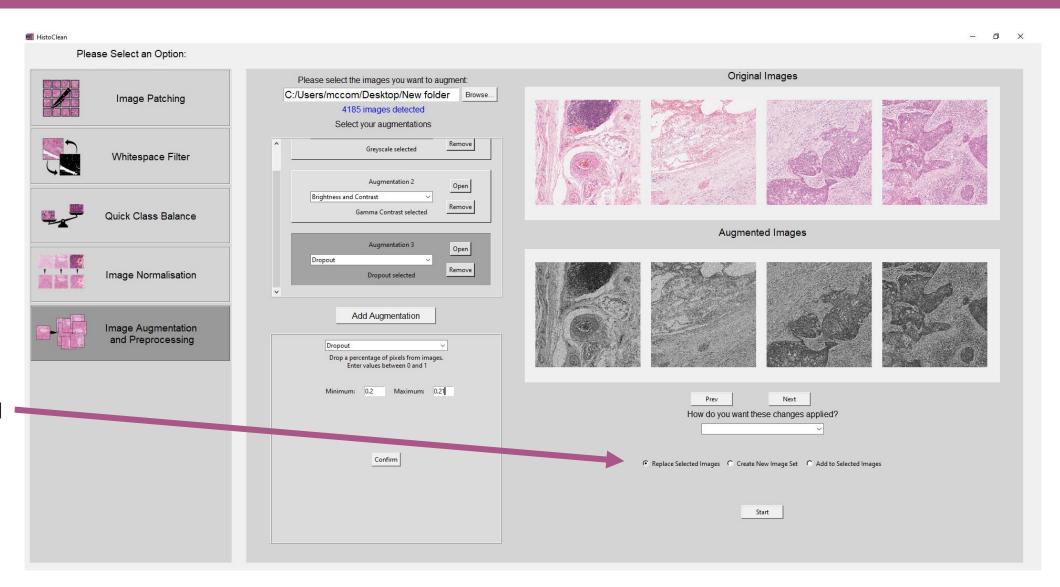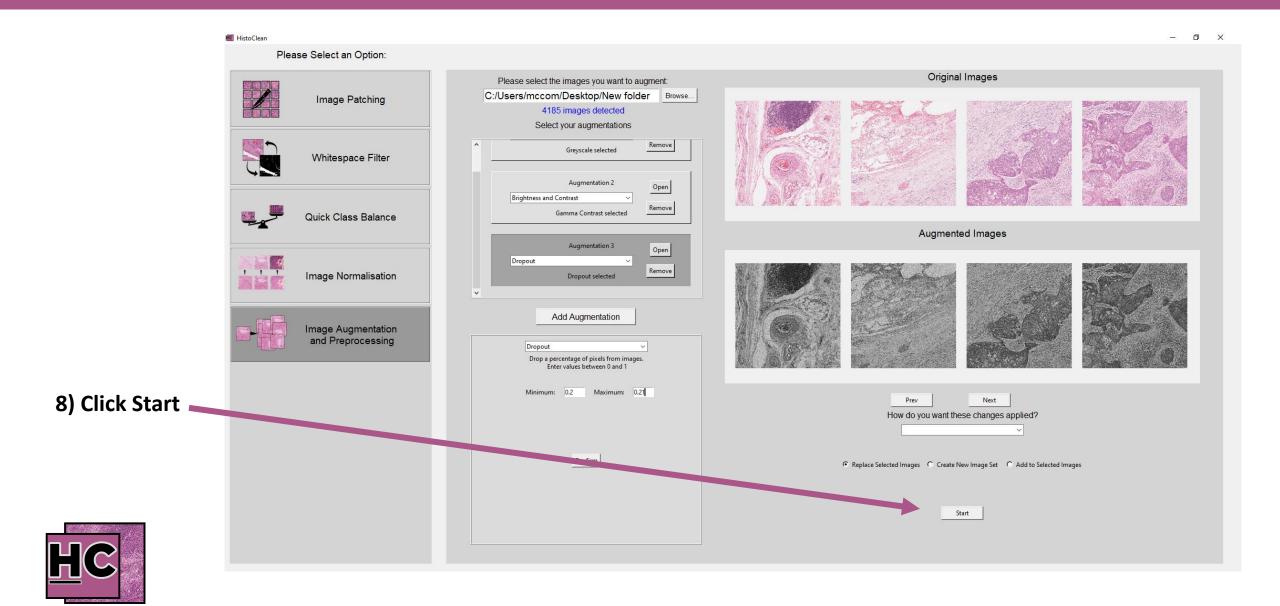**5) Save each effect by clicking the "confirm" button.**

5) Image Augmentation and Pre-processing

6) Select if you want these changes applied to all images, a certain number of images, or a random number of images

7) Decide whether to replace the current images, create a completely new image set, or add to the current images.

5) Image Augmentation and Pre-processing

8) Click Start

The Output folder structure follows that of the input structure.