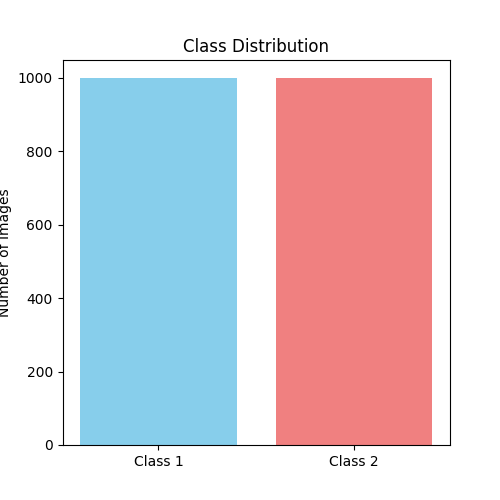
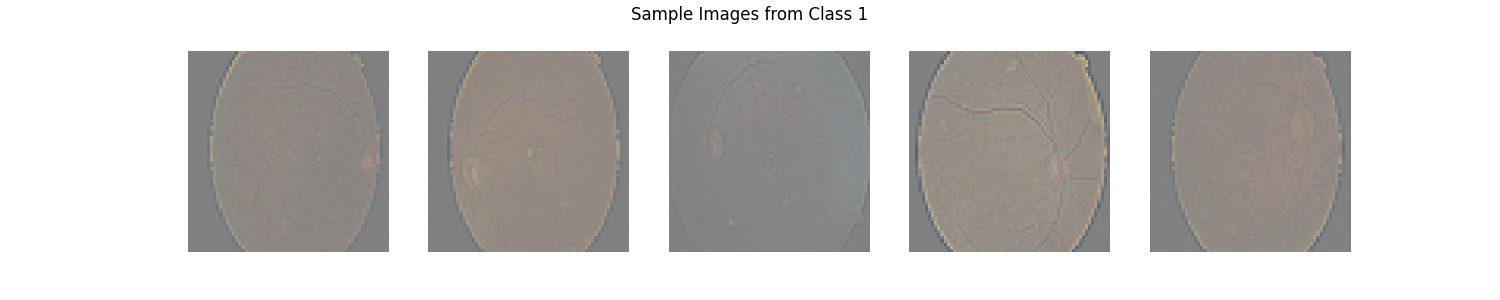
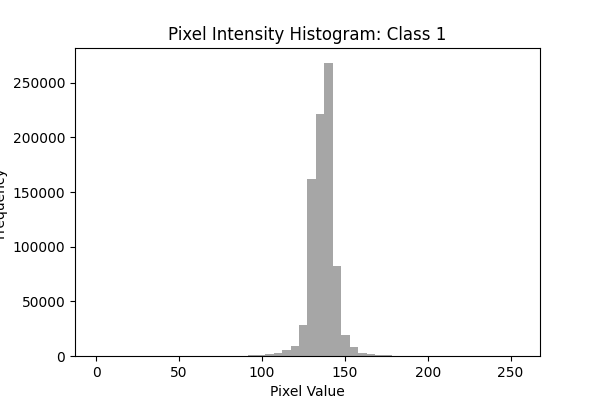
**Class distribution plot**

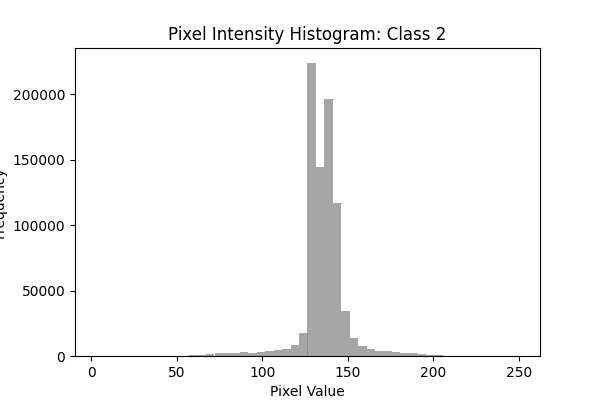
We assessed the distribution of images across the two classes to ensure that the dataset was balanced. The distribution plot (Figure X) indicates approximately 999 images in each class, suggesting an almost perfectly balanced dataset. Balanced datasets are required to avoid bias when training models, and they achieve fair and consistent classification performance across the categories**.**

**Sample images**  
  
To gain some qualitative insights into the dataset, we visualized a subset of sample images from Class 1 (Figure X). This allows us to assess intra-class variation, general visual attributes, and likely noise/image artifacts in the dataset. This visualization helps to understand the general texture, colour, and shape features that the model will learn to discriminate between.



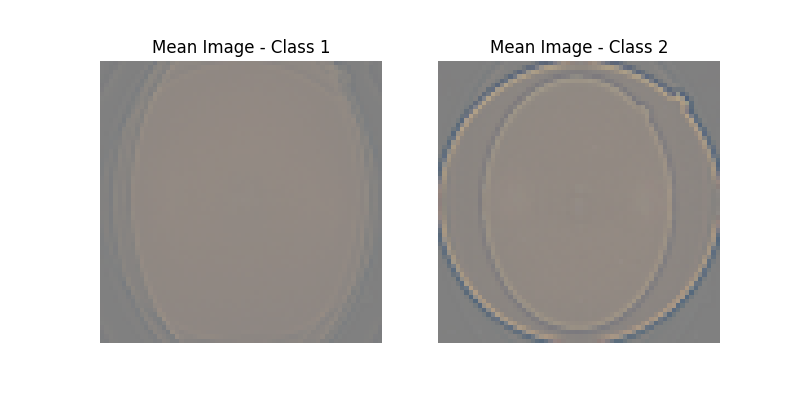
In the same way, we provide a selection of sample images from Class 2 in Figure X. Reviewing these images provides a qualitative understanding of the unique visual characteristics of this class, which in combination with Class 1, allows us to evaluate both inter-class differences and intra-class differences. If the two classes can be qualitatively determined to be different enough for supervised learning purposes, we have sufficient spectral variation across both classes.

**Histogram of pixel intensity**

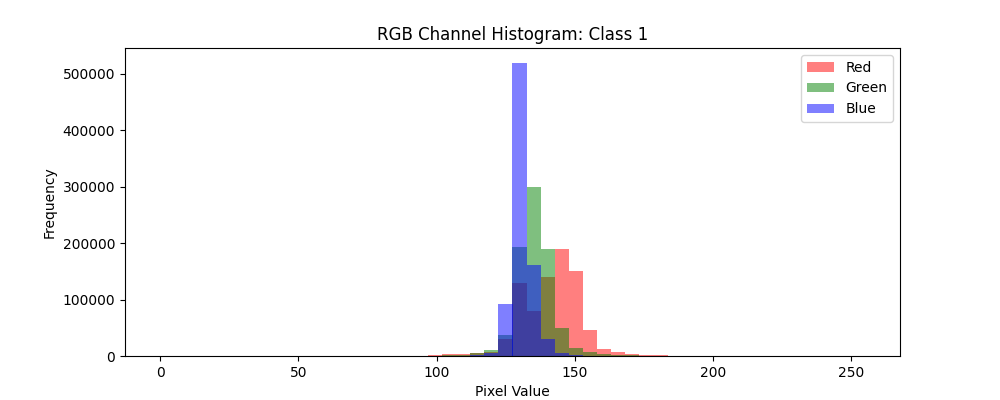
**Class1:** We examined the distribution of pixel intensity values for Class 1 images with a grayscale histogram (Figure X), which indicates the spread of brightness levels for the entire class. The histogram shows a broad range of pixel intensity values, a reflection of differing levels of illumination and some variety in contrast, both of which could influence the ability of the model to generalize.

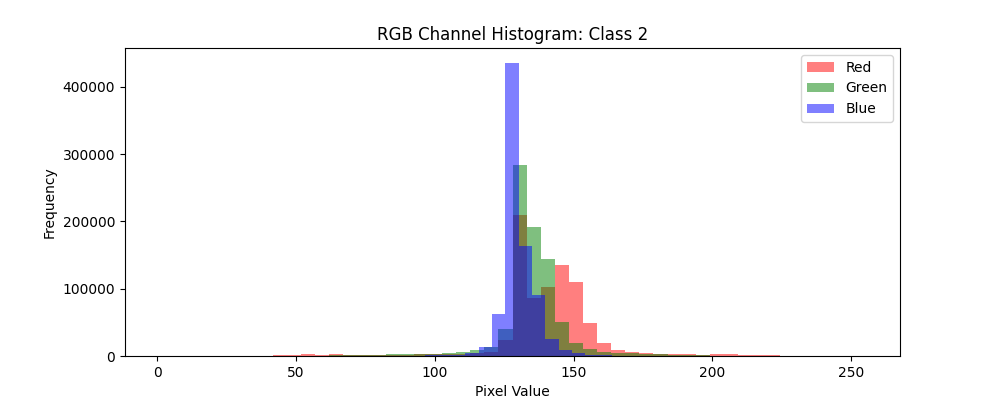
Class 2: In contrast, the grayscale histogram of Class 2 shows a different distribution of pixel intensity values (Figure X). This tells us that Class 2 has variations of brightness and contrast in pixel intensity specific to its class. This information helps the user spot the possible need for preprocessing (i.e. normalization or contrast transformation) and gives the user some understanding of the inherent differences that can be used by the model correctly classified instances.

**Average (mean) image**

To further visually summarize this information for each class, we computed the mean image, by averaging all images in a class (Figure X). The mean images reveal common patterns, prevalent colors, and general shapes for each category. The differences in the mean images suggest the classes contain learnable features by convolutional neural networks. A file of this nature can be completed as a further indicator of dataset separability.

**RGB channel histogram  
Class 1:**  
We depicted the RGB channel histograms to investigate color distribution in the Class 1 images (Figure X). The histograms identify the proportion of pixels for each of the three channels (red, green, and blue) for the images and identify the colors that had the highest proportion (thus the most dominant color) and helps identifying whether there may be a color signature for the class-specific colors. This is important to assess in color-based classification tasks and whether the color-related channels need normalizing.



  
**Class 2:**  
  
Similarly, the RGB histograms for Class 2 (Figure X) provide some shape to the unique color distributions that were not evident in Class 1. The peaks in the various channels compared to Class 1 indicate that color can be centred on as a viable distinguishing feature. Understanding these differences will aid in preparing a model and pre-processing.