

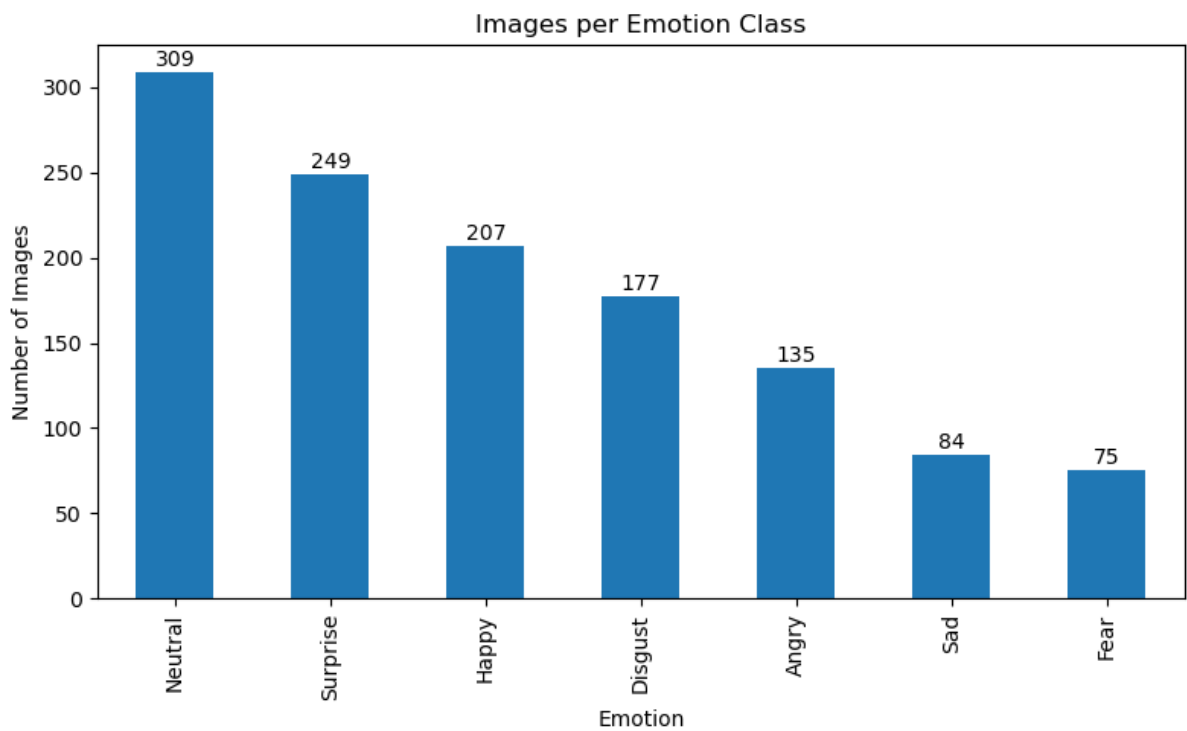
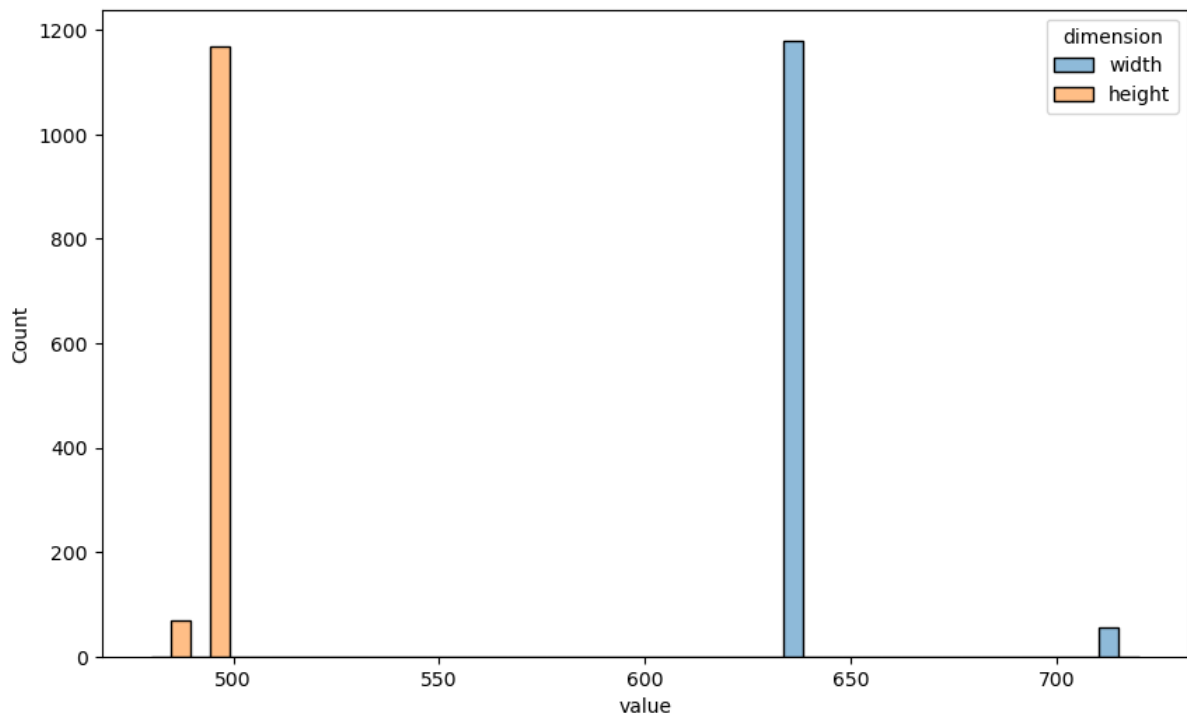
Experimental Data Analysis: CK + dataset

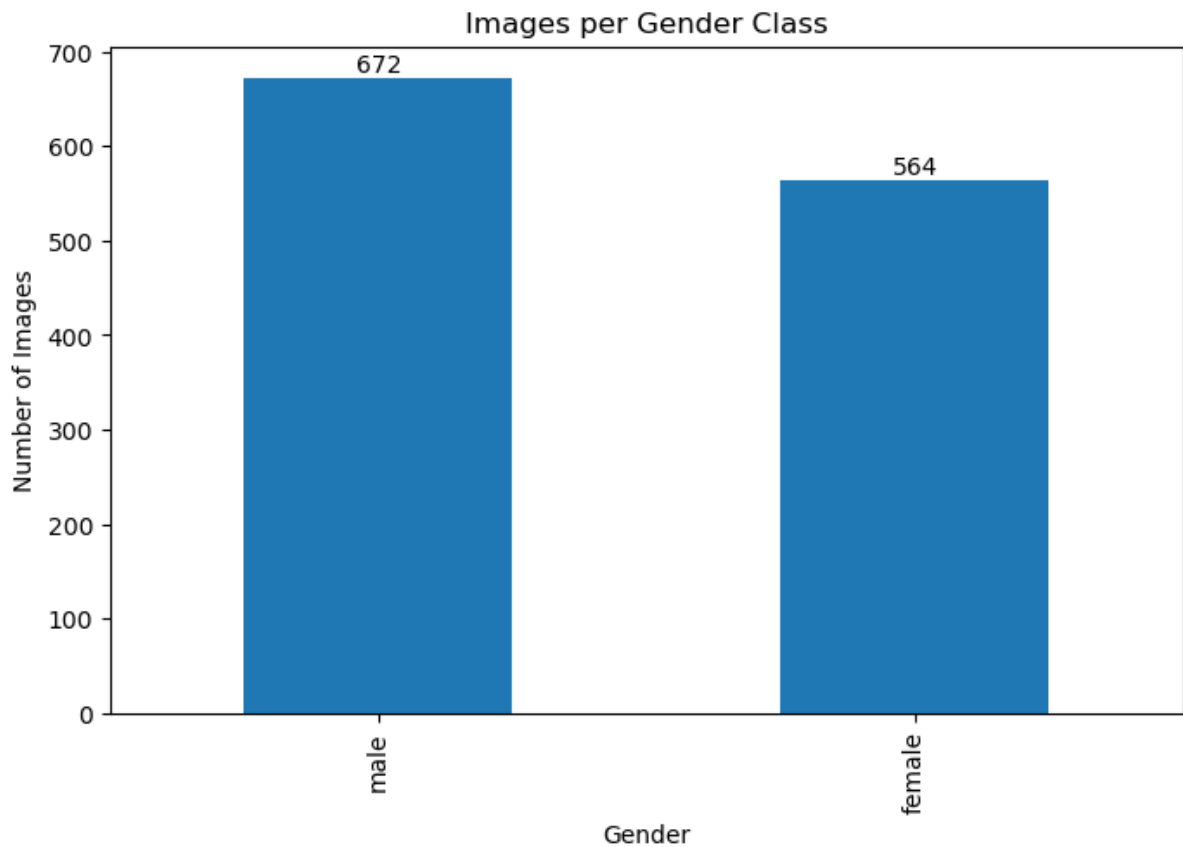
Plotting random images, as we can see they are in different format, some are RGB, some grayscale.



In our previous discussion, we were surprised to notice that each person in the dataset appears in three different images (sequence). After doing some research and reading a few papers, my hypothesis is that recording a person's facial expressions as a short video (something like 3-5 frames), rather than taking a single still photo is much more informative when trying to capture their emotions. Subtle, almost unnoticeable facial movements, such as eye changes, lip movement, or cheek tension, often reveal emotional states more accurately. Video allows us to observe more expressions, which could significantly improve emotion recognition compared to analyzing a single static frame.

Statistics





Summary:

The difference in amount of samples across classes is quite big. It might be worth considering techniques to balance this difference — such as downsampling the larger classes or applying other class balancing methods — and carefully deciding which labels we actually want to fine-tune our model on. The difference in gender is not that big, I wouldn't worry about that one.