How to Handle Imbalanced Data:

To balance the dataset and make sure rare cancer types are adequately learned, the following strategies need to be considered:

1) Data Augmentation: Generate more images for underrepresented classes (like vascular lesions) using:
   - Rotation, flipping, zooming
   - Color jittering, brightness/contrast adjustment
   - Advanced method: Use GANs (Generative Adversarial Networks) to synthesize realistic images
2) Class Re-weighting in Loss Function:
   - Tell the model: “Make more effort to correctly classify rare cancers” by assigning higher weight in the loss function to underrepresented classes.
   - Common technique: Weighted Cross-Entropy Loss
3) Resampling (Under/Over-sampling)
   - Oversample rare classes by duplicating or augmenting images
   - Undersample common classes (only use a subset of melanocytic nevi images)
4) Focal Loss
   - Modify the loss function so the model focuses more on hard-to-classify examples, often from the minority classes.