Supplementary Materials

6 Detailed Dataset Information

We illustrate the class distribution across each dataset in Fig. 4 and Fig. 5, highlighting the significant imbalance. Additionally, we present representative samples from each class in Fig. 6.

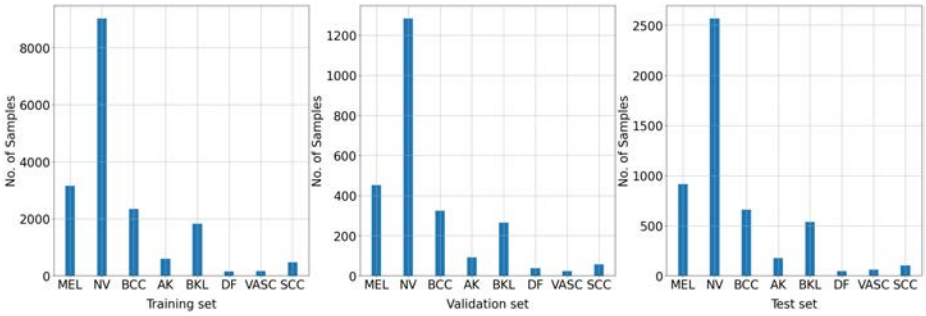


Fig. 4: Distribution of the classes in ISIC-2019 Dataset

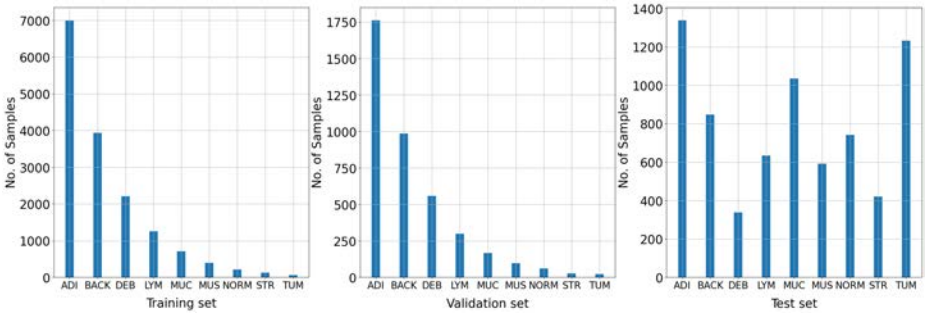
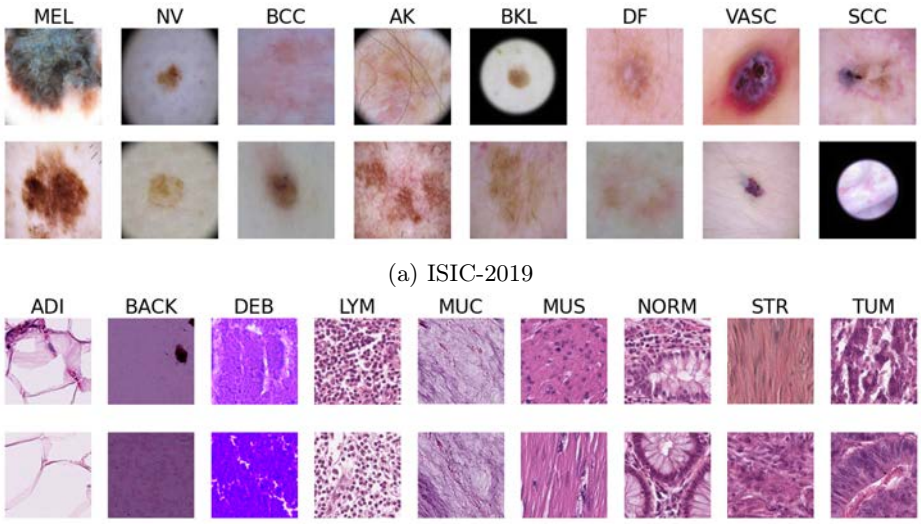


Fig. 5: Distribution of the classes in long-tailed NCT-CRC-HE-100K Dataset

7 Hyperparameter Analysis: Mix Ratio (m)

In Fig. 7, we compare the impact of the mix ratio (m) in Co-teaching VOG, using the macro-averaged test F1-score obtained after training with noisy labels

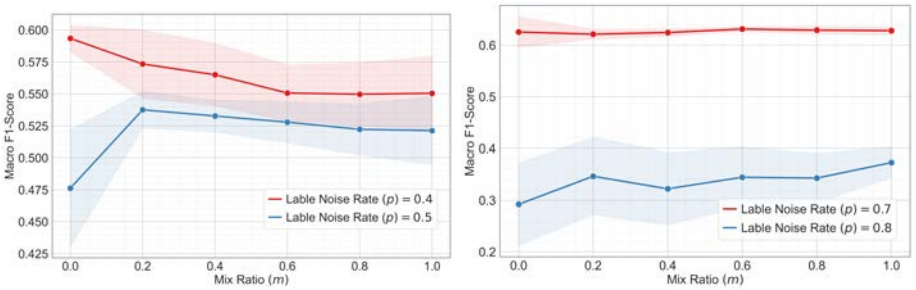


(a) ISIC-2019

(b) Long-tailed NCT-CRC-HE-100K

Fig. 6: Some examples of images of the two datasets

in the initial LNL phase. These results indicate that this hyperparameter differs across datasets and can vary with label noise (p).



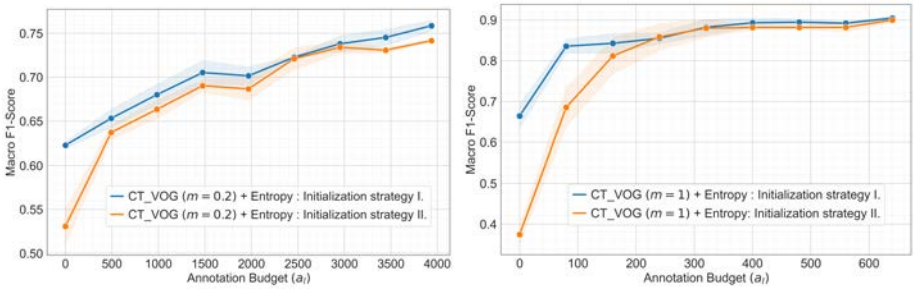
(a) ISIC-2019 (b) LT NCT-CRC-HE-100K

Fig. 7: Hyperparemeter study of mix ratio (m) in two datasets, when training Co-teaching VOG at the first phase.

7.1 Base Model Initialization Strategy

There are two strategies to initialize the base model in the first phase before the active label cleaning round begins: I. either use the model trained on a noisy

dataset using Co-teaching VOG (similar to [2]) or II. use the samples selected by Co-teaching VOG as clean labels to train a new model using standard cross-entropy loss. In Fig. 8, we compared these strategies and observed that separately training the model using standard cross-entropy with only the samples identified by Co-teaching VOG as clean labels improved the initial performance the most. We argue that by segregating the noisy samples from an early stage, we reduce the possibility of model distortion due to noisy labels. Therefore, we adopted strategy I. for Co-teaching VOG, as reported in the Results section.



(a) ISIC-2019 ($p = 0.5$) (b) LT NCT-CRC-HE-100K ($p = 0.8$)

Fig. 8: Comparison of test performance (macro-averaged F1-score) using two initialization strategies across two label noise rates (p) in two datasets. Initialization strategy I. refer to retraining the model from scratch using cross-entropy with only the clean labels identified by Co-teaching VOG. Initialization strategy II. refers to directly utilizing the model trained with Co-teaching VOG on noisy labels.