ISIC 2017 Challenge Task 3: Melanoma Classification

Methodology

This project addresses binary skin lesion classification for Task 3 of the ISIC 2017 Challenge. The objective is to distinguish melanoma from non-melanoma lesions using dermoscopic images.

We employed a transfer learning strategy using the EfficientNet-B3 architecture pre-trained on ImageNet. The model was fine-tuned using a class-balanced, augmentation-rich pipeline and evaluated on the ISIC 2017 dataset.

The dataset includes 500 training images, 150 validation images, and 600 test images. To increase generalization, extensive data augmentation was applied to the training images, including resizing to 224×224, random horizontal/vertical flips, random rotations, color jitter, and random erasing. Validation and test images were only resized and normalized using ImageNet statistics.

To address class imbalance, we used a weighted random sampler and a class-weighted cross-entropy loss. The model was trained using the Adam optimizer with an initial learning rate of 1×10^{-4} and a weight decay of the same magnitude. Gradient clipping was used to prevent exploding gradients. A StepLR scheduler reduced the learning rate by a factor of 0.1 every 5 epochs. Training was capped at 20 epochs with early stopping (patience = 3) based on validation AUC.

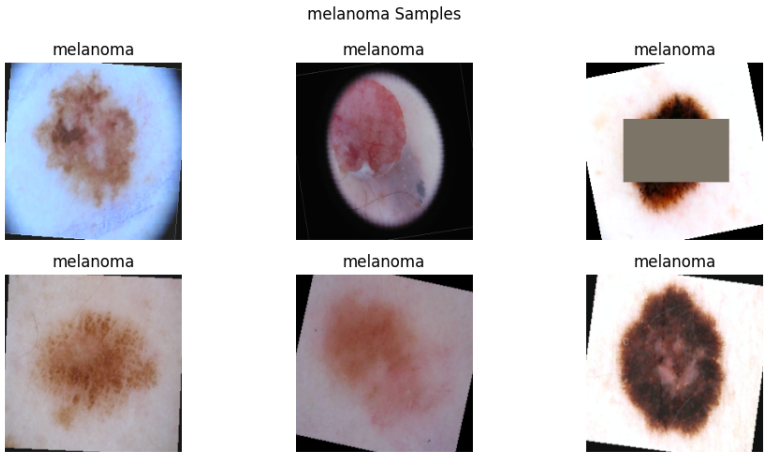

Figure 1: Sample melanoma images from the training set.

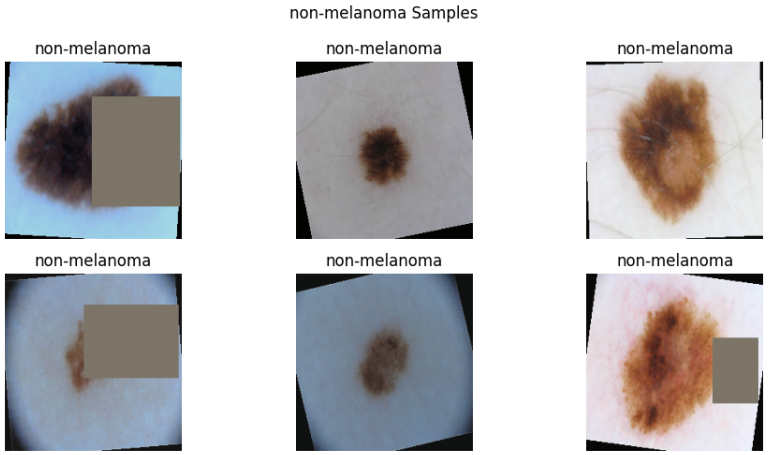

Figure 2: Sample non-melanoma images from the training set.

Model Architecture and Optimization

We selected EfficientNet-B3 for its balance between performance and computational efficiency. The classifier head was replaced to output two logits corresponding to the binary labels. Optimization was performed using Adam, a widely used adaptive learning rate method. Learning rate scheduling helped refine convergence, and early stopping prevented overfitting.

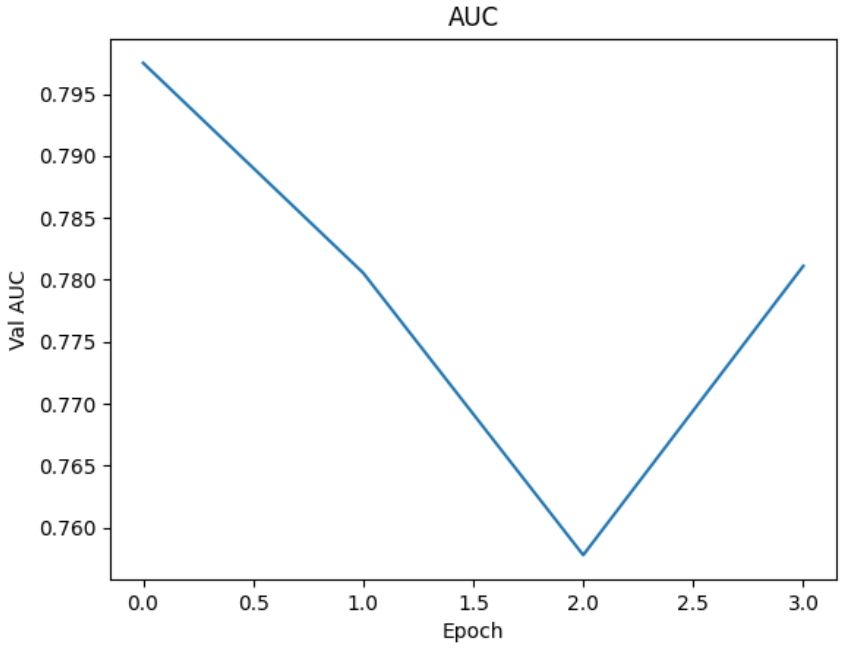

Figure 3: Validation AUC across training epochs.

Evaluation Results

The best model achieved a test AUC of **0.707**. Table 1 summarizes the performance on the held-out test set using a fixed threshold of 0.5. The model achieved 0.69 accuracy, with stronger performance in identifying non-melanoma cases, but a promising recall (sensitivity) of 0.62 for melanoma.

Table 1: Classification Report on the Test Set

Class	**Precision**	**Recall**	**F1-score**	**Support**
Non-Melanoma	0.88	0.71	0.79	483
Melanoma	0.34	0.62	0.44	117
Accuracy			0.69	
Macro Avg	0.61	0.66	0.61	600
Weighted Avg	0.78	0.69	0.72	600

Conclusion

The trained EfficientNet-B3 model demonstrated promising performance on the ISIC 2017 test set, particularly in detecting melanoma cases—a critical factor in clinical screening. While the recall for melanoma was satisfactory, the relatively low precision indicates a tendency toward false positives. The model was trained on a subset of 500 images from the full set of 2,000 available training samples, which likely constrained its overall performance. Future improvements could include utilizing the complete dataset, applying threshold tuning, ensembling multiple models, or incorporating domain-specific augmentation techniques to improve both robustness and precision.