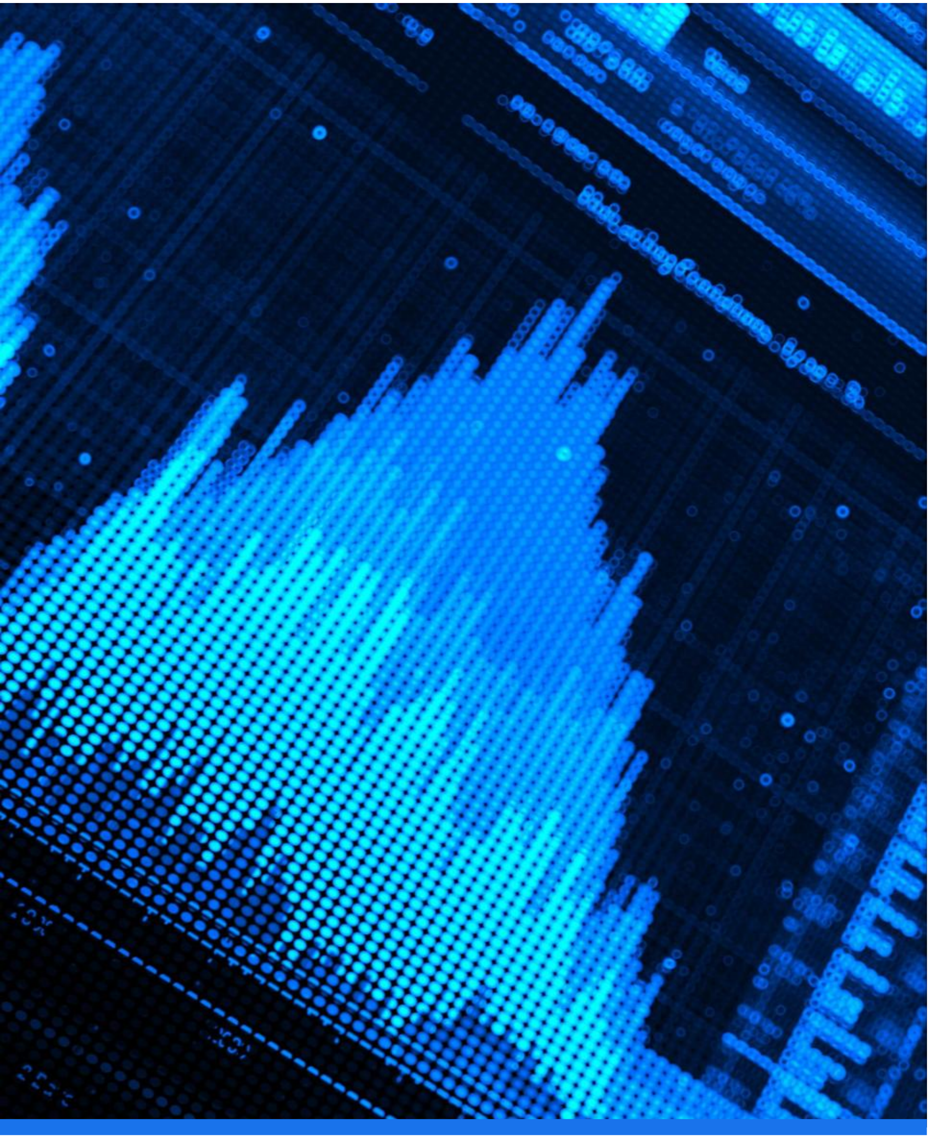
Classification Model for Lesion Images on the HAM10000 Dataset

CONOR HUH, MRIDUL JAIN, VISHAL SAXENA, LYNNE WANG

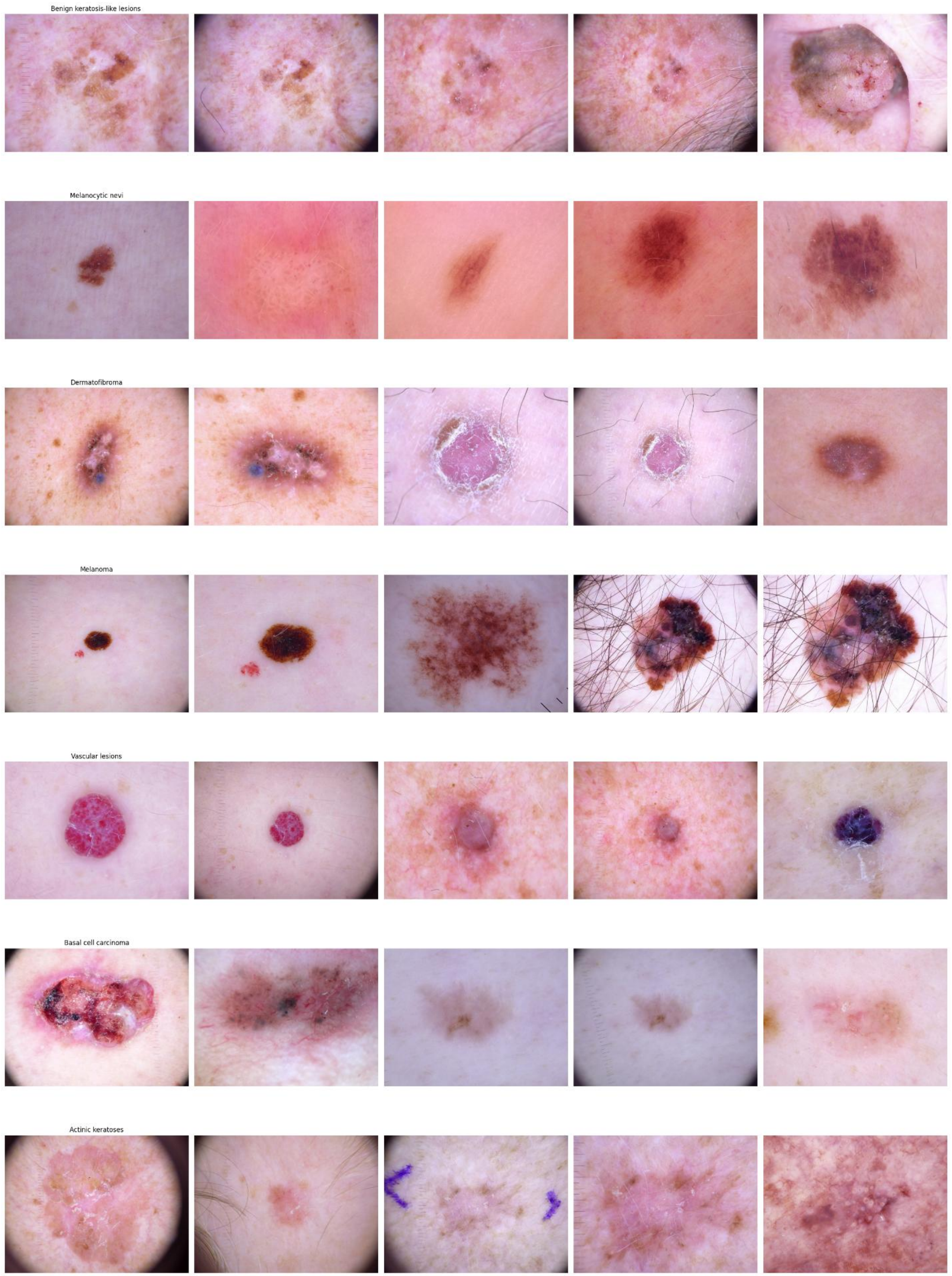
MIDS 281 COMPUTER VISION, UC BERKELEY



Presentation Roadmap

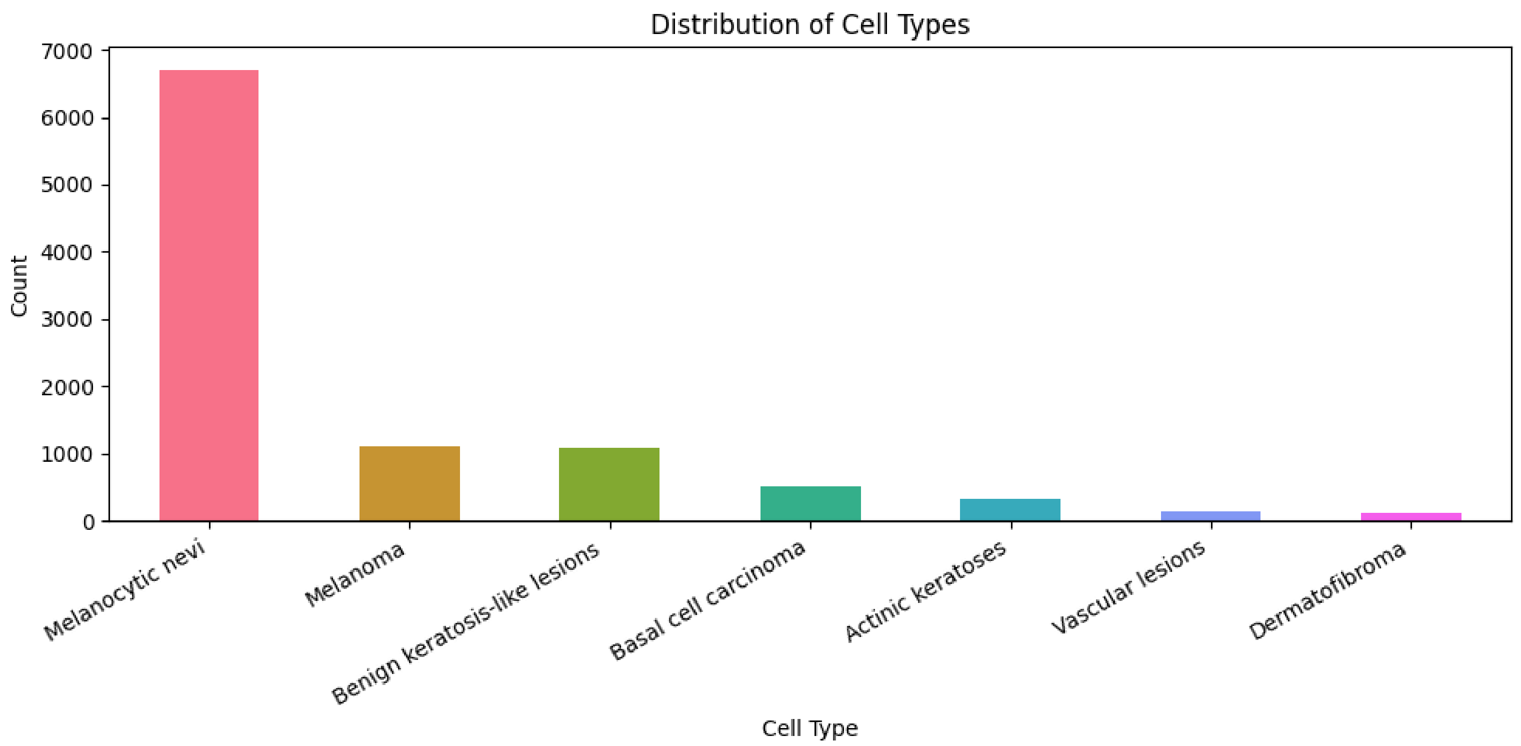
- Addressing Class Imbalance in the HAM Dataset: Upsampling Techniques
- Feature Extraction: Simple and Complex Approaches for Lesion Classification
- Feature Selection Rationale and Interpretation
- Classification Methods and Performance Analysis
- Ensuring Generalizability: Data Splitting, Hyperparameter Search, and Evaluation
- Balancing Efficiency and Accuracy: Comparative Analysis of Feature-Classifier Combinations

Overview of the HAM Lesion Dataset



Addressing Class Imbalance in the HAM Dataset:

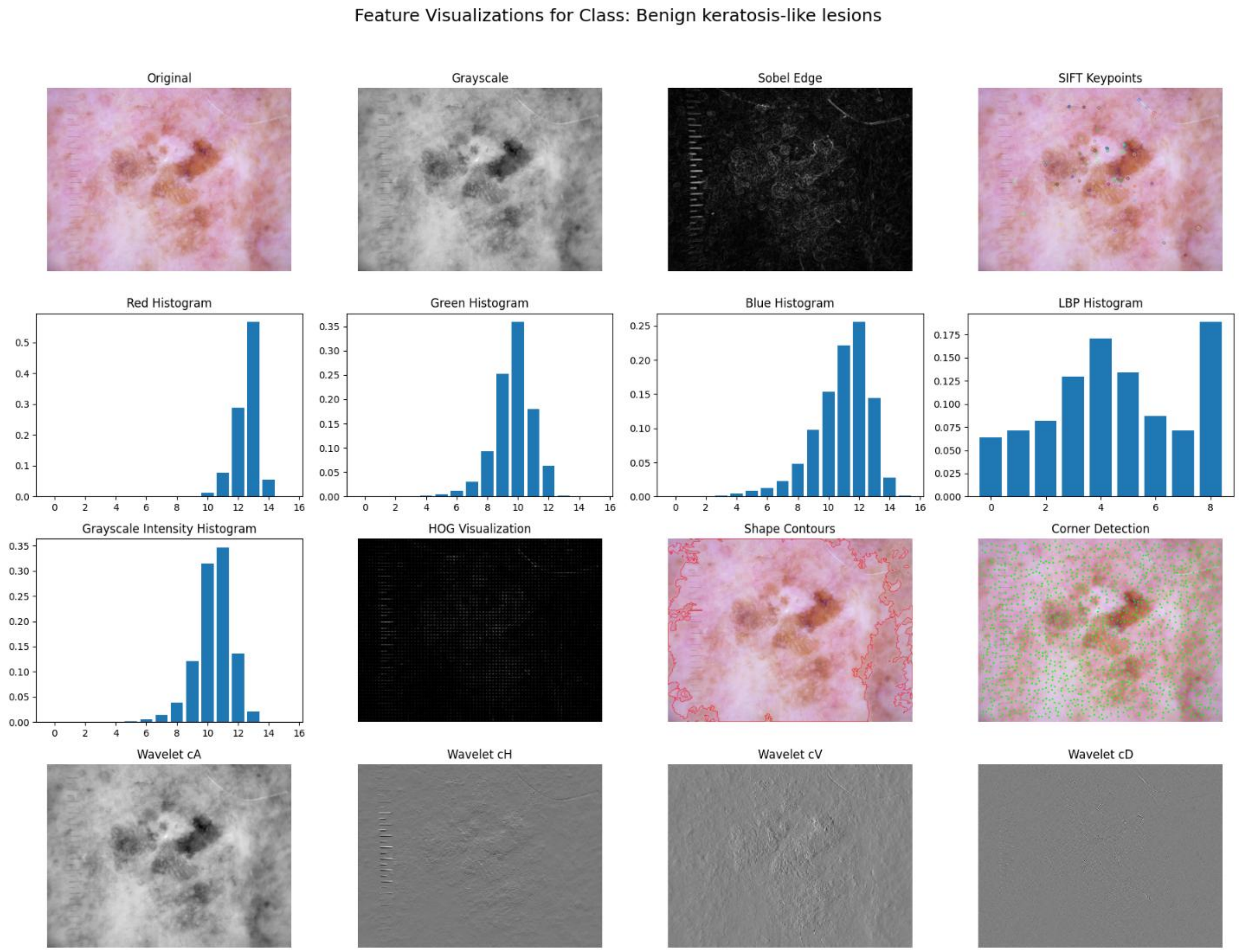
1. Oversampling
2. Class Weighting

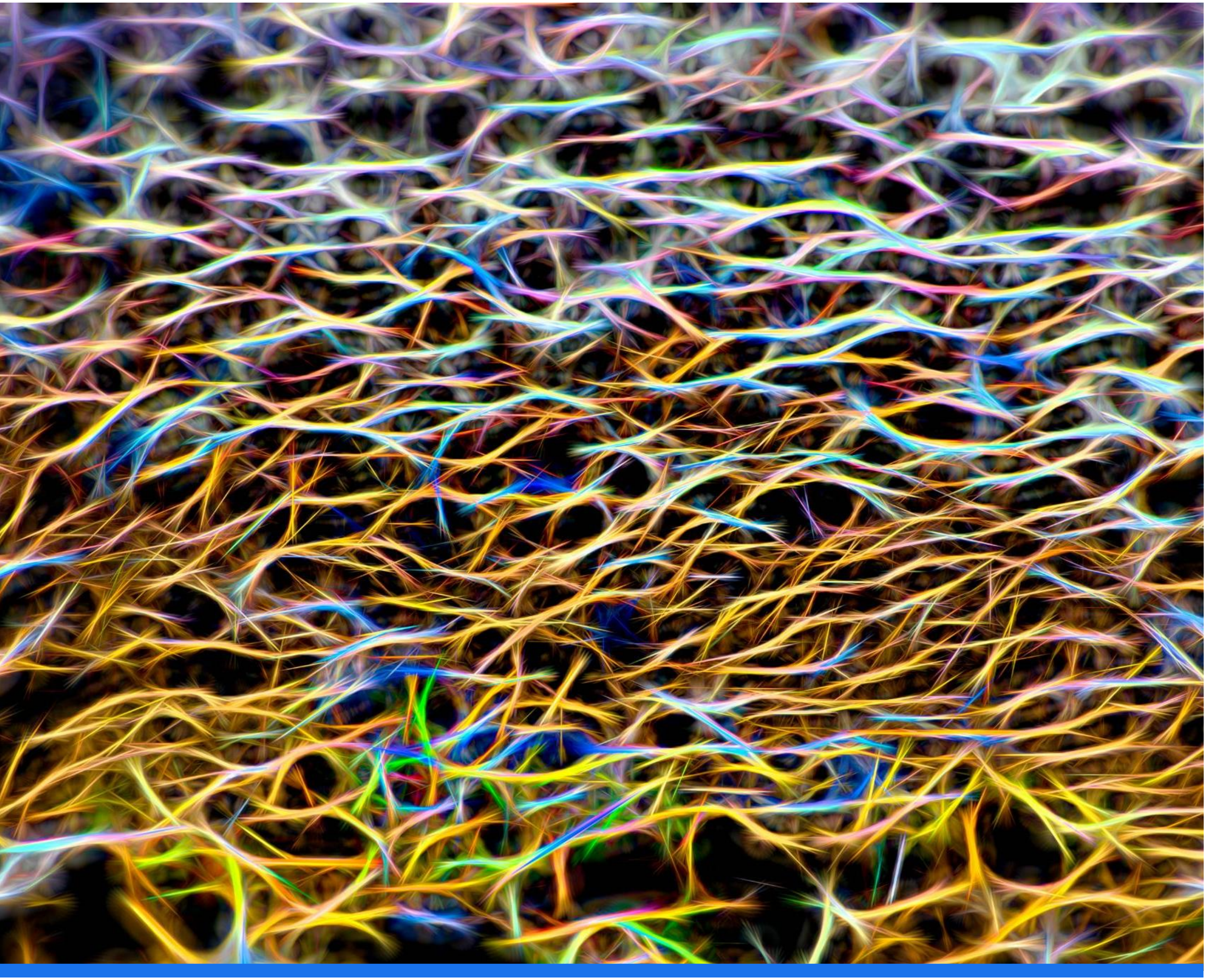


Simple Feature Extraction

Feature Type	Description	Feature Dimension
HSV	Normalized 3D histogram of hue, saturation, and value channels	1024
HOG	Gradient-based edge orientation features in grayscale	8100
LBP	Histogram of uniform local binary patterns (rotation invariant)	59
Sobel	Descriptive stats on edge magnitude (mean, std, skew, kurtosis, percentiles, entropy)	8
Wavelet	Approximation mean + energy from detail coefficients	4
Color	Mean & std of RGB + normalized 3D RGB histogram	518
GLCM	Texture features via gray-level co-occurrence matrix	6
Gabor	Mean & std dev from multiple filtered responses (6 orientations × 4 scales)	48
Shape	Area, perimeter, compactness of the largest contour	3
Corner	Number of corners detected using Shi-Tomasi method	1
SIFT	SIFT keypoint descriptors flattened to fixed length vector (50 x128)	6400
Intensity	Statistical moments: mean, std, skewness, kurtosis of grayscale intensity	4

Visualization of Simple Features

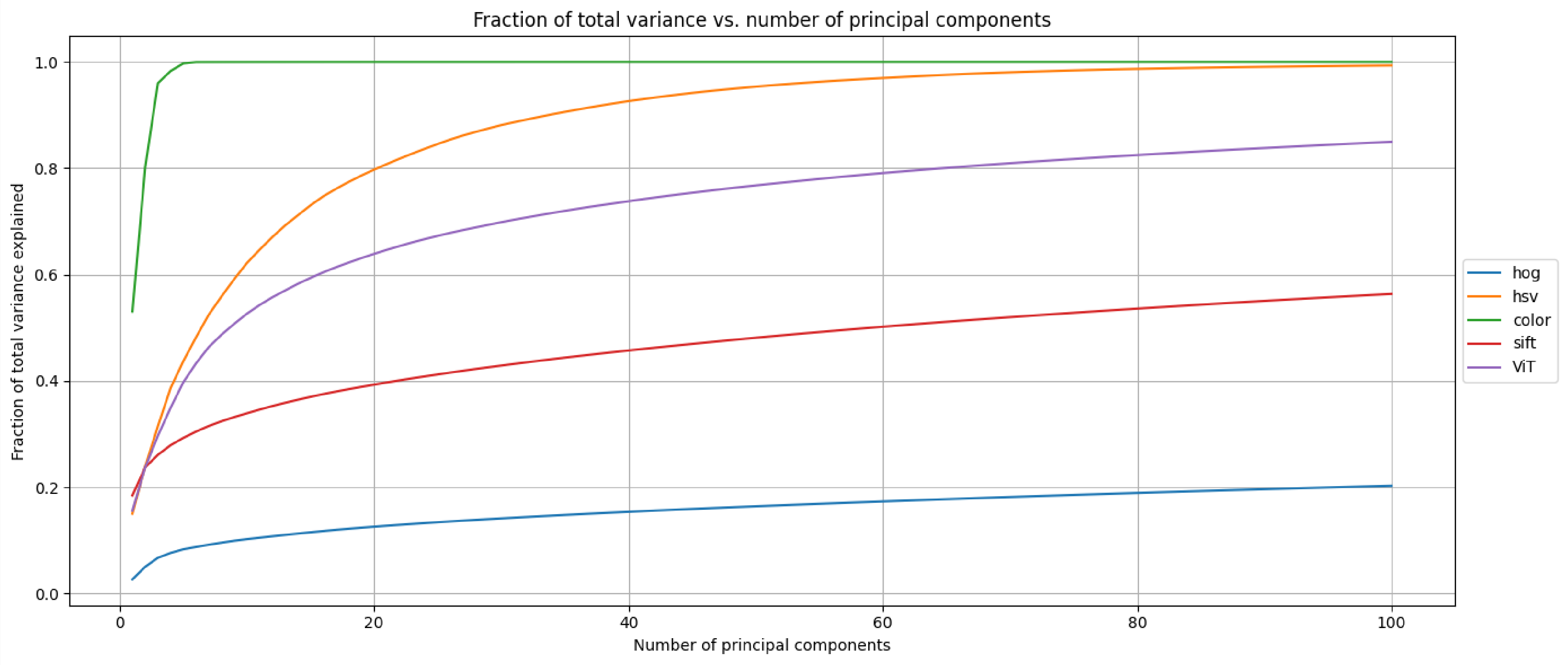




Complex Feature Extraction:

1. Vit
2. Clip
3. Resnet

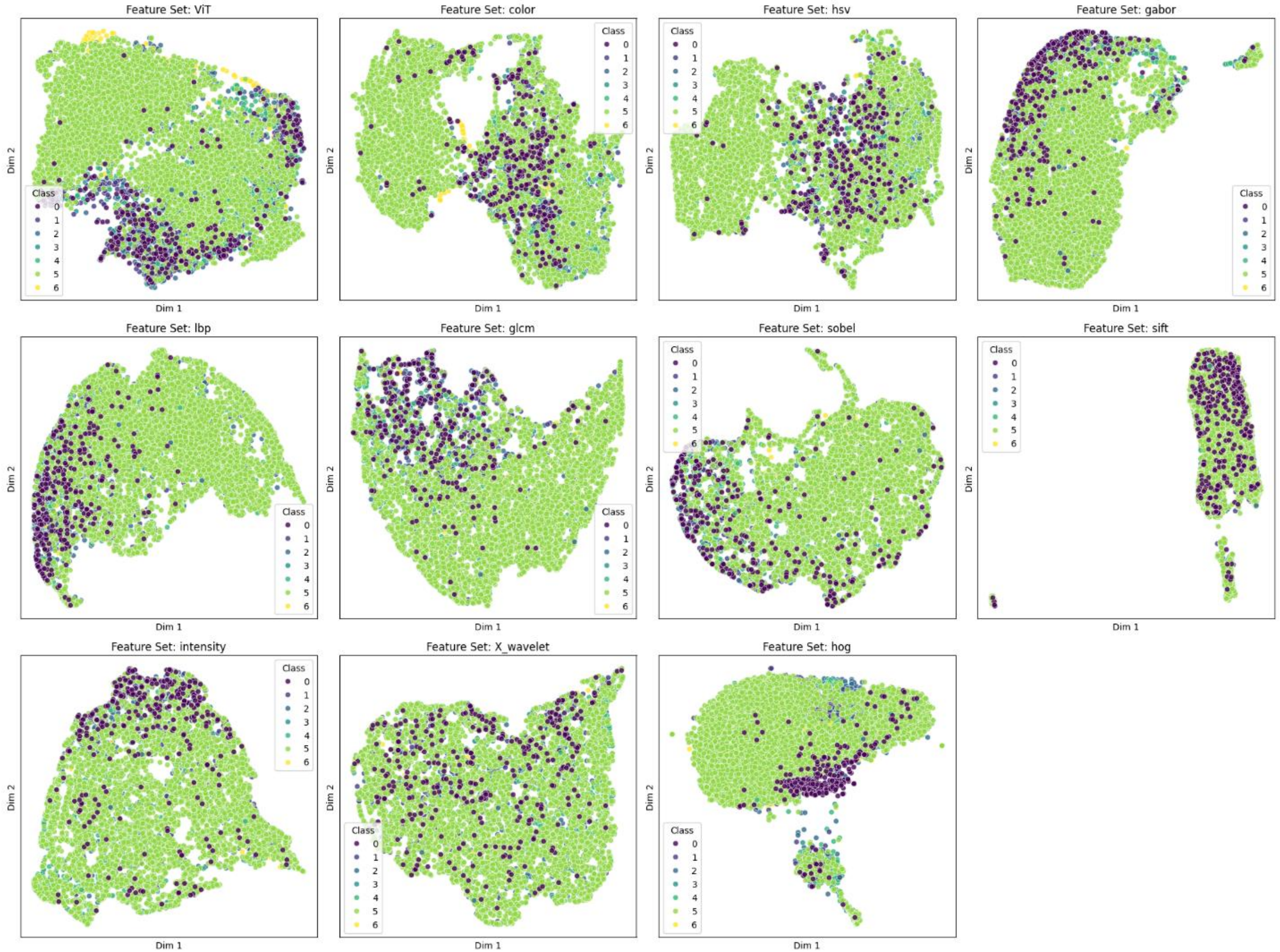
PCA of High Dimensional Features



AUC-ROC of Each Feature

	mean_auc_roc	akiec	bcc	bkl	df	mel	nv	vasc
ViT	0.918807	0.940296	0.946014	0.887043	0.913504	0.85162	0.908344	0.984825
color	0.905113	0.903213	0.915098	0.86108	0.885808	0.881939	0.899282	0.989368
hsv	0.900603	0.896002	0.910165	0.853191	0.867202	0.887739	0.897905	0.992014
gabor	0.782414	0.881289	0.810262	0.734217	0.734684	0.777433	0.814387	0.724627
lbp	0.741303	0.886965	0.730321	0.660882	0.672998	0.739452	0.801442	0.697058
glcm	0.72045	0.829348	0.74523	0.680216	0.706693	0.730609	0.797681	0.553369
sobel	0.709296	0.79191	0.722999	0.631536	0.59653	0.717012	0.76058	0.744508
sift	0.701086	0.805895	0.732431	0.676648	0.618868	0.761287	0.754935	0.557536
intensity	0.684763	0.7843	0.750177	0.642779	0.616066	0.689608	0.733427	0.576982
X_wavelet	0.631446	0.746668	0.615292	0.554721	0.553793	0.676672	0.673983	0.598994
hog	0.618517	0.747637	0.566517	0.599901	0.592902	0.557032	0.624516	0.641112
shape	0.614213	0.708403	0.657218	0.614838	0.527211	0.541862	0.677848	0.572112
corners	0.507599	0.500864	0.503842	0.501728	0.514527	0.5002	0.500283	0.531749

2D Visualization of Features



Training and Evaluation Strategies

K-Fold
+
Oversampling or Class-Weighting

- We first performed a stratified 80/20 train-test split, ensuring class distribution was preserved. The 20% test set was held out entirely for final evaluation.
- The 80% training set was then further split (80/20) into training and validation subsets, and used with 5-fold Stratified K-Fold Cross-Validation for robust hyperparameter tuning.
- We used Bayesian optimization (BayesSearchCV) with 5-fold cross-validation to identify optimal hyperparameters. For each hyperparameter candidate, five models were trained and evaluated across folds, and the average performance (macro F1 or AUC-ROC) was used to guide optimization.
- After selecting the best hyperparameters, a final model was retrained on the entire 80% training set, incorporating the same oversampling or class-weighting strategy.
- Threshold tuning was performed on the validation set to improve class-wise decision boundaries and enhance macro F1 performance on the multiclass task.

Model Performance

F1 Optimized Models w/ Oversampling							
Model	Train Time (s)	Inference Time (s)	Best CV F1-Macro	Test Set AUC-ROC	Test Accuracy	Test Macro F1-Score	Best Hyperparameters
XGB*	2364.96	0.11	0.6752	0.9699	0.8509	0.7407	lr=0.29, max_depth=3, n_est=491
LS	347.99	0.03	0.6556	0.9534	0.7902	0.6712	C=2.07, penalty='l1'
RF	347.47	0.31	0.5585	0.9347	0.6983	0.6119	max_depth=10, n_bins=19, n_est=500
KNN	105.28	0.17	0.3423	0.6654	0.613	0.3496	n_neighbors=3, p=2

AUC-ROC Optimized Models w/ Oversampling							
Model	Train Time (s)	Inference Time (s)	Best CV AUC-ROC	Test Set AUC-ROC	Test Accuracy	Test Macro F1-Score	Best Hyperparameters
XGB	707.4	0.11	0.9566	0.9649	0.8268	0.6948	lr=0.30, max_depth=3, n_est=412
RF	86.32	0.3	0.9435	0.9514	0.7907	0.6229	max_depth=45, n_bins=18, n_est=500
LR	347.99	0.03	0.9385	0.9534	0.7987	0.6761	C=0.69, penalty='l1'
KNN	6.96	0.07	0.7335	0.7221	0.5452	0.3142	n_neighbors=30, p=2

Model Performance

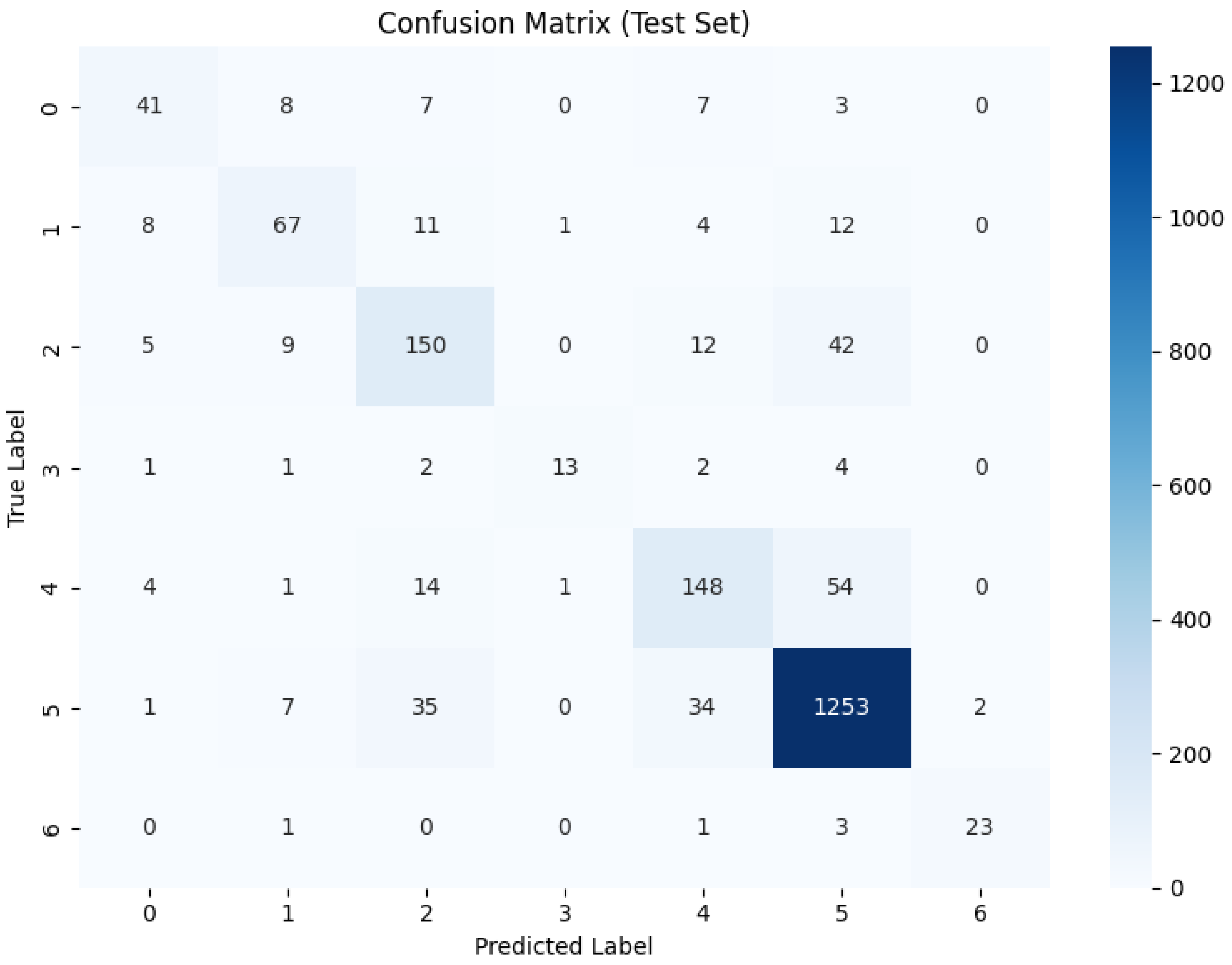
F1 Optimized Models w/ Class Balancing							
Model	Train Time (s)	Inference Time (s)	Best CV F1-Macro	Test Set AUC-ROC	Test Accuracy	Test Macro F1-Score	Best Hyperparameters
XGB	2038.62	0.05	0.6244	N/A	0.8333	0.6887	lr=0.16, max_depth=5, n_est=319
LR	1627.75	0.02	0.6369	N/A	0.8047	0.6742	C=62.71, penalty='l1'
RF	302.21	0.43	0.4007	N/A	0.7585	0.4048	max_depth=33, n_bins=101, n_est=212
KNN	84.33	0.09	0.3142	N/A	0.7083	0.3421	n_neighbors=15, p=2

AUC-ROC Optimized Models w/ Class Balancing							
Model	Train Time (s)	Inference Time (s)	Best CV AUC-ROC	Test Set AUC-ROC	Test Accuracy	Test Macro F1-Score	Best Hyperparameters
XGB	151.56	0.05	0.9499	0.9587	0.8228	0.6638	lr=0.17, max_depth=14, n_est=187
LS***	76.91	0.02	0.9457	0.9472	0.7942	0.5898	C=105.76, penalty='l2'
RF	21.97	0.27	0.9352	0.9504	0.7952	0.5842	max_depth=23, n_bins=95, n_est=470
KNN**	1.14	0.04	0.827	0.8194	0.7063	0.3261	n_neighbors=26, p=2

Our Best Model

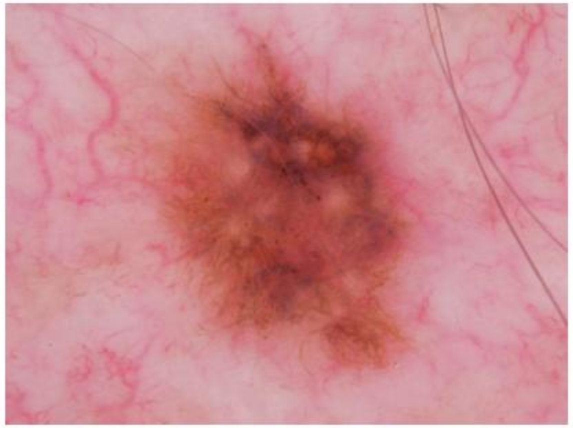
Class	Precision	Recall	F1-Score	Support
0 (Actinic keratoses)	0.683333	0.621212	0.650794	66
1 (Basal cell carcinoma)	0.712766	0.650485	0.680203	103
2 (Benign keratosis-like lesions)	0.684932	0.688073	0.686499	218
3 (Dermatofibroma)	0.866667	0.565217	0.684211	23
4 (Melanoma)	0.711538	0.666667	0.688372	222
5 (Melanocytic nevi)	0.913931	0.940691	0.927118	1332
6 (Vascular lesions)	0.92	0.821429	0.867925	28
accuracy	0.850904	0.850904	0.850904	
macro avg	0.784738	0.707682	0.740732	1992
weighted avg	0.847812	0.850904	0.848619	1992

Our Best Model

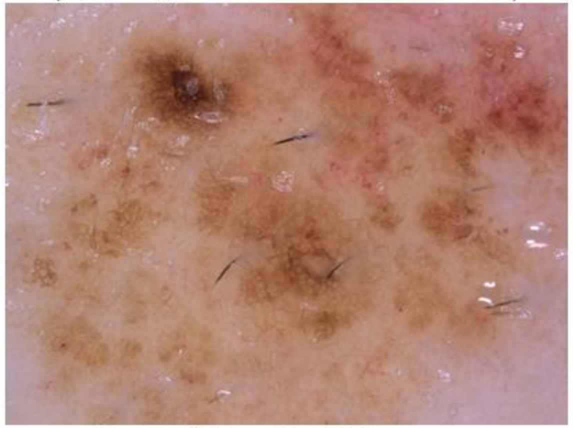


Case Analysis: Misclassification Between Class 4 and Class 5

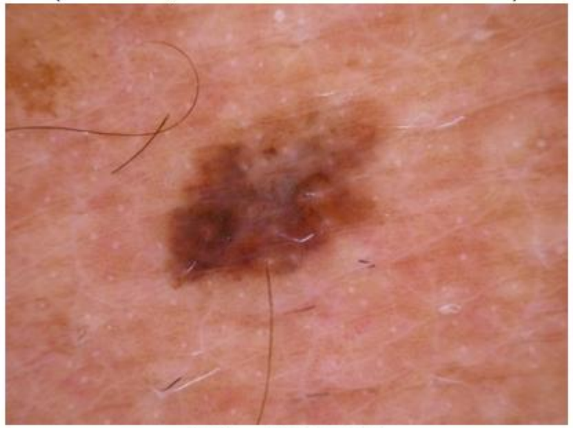
ISIC_0029272
(Class 4, misclassified as Class 5)



ISIC_0026930
(Class 4, misclassified as Class 5)



ISIC_0030771
(Class 5, misclassified as Class 4)



ISIC_0033266
(Class 5, misclassified as Class 4)

Future Work:

Feature Refinement and Expansion: revisiting shape, boundary, color, and edge-based features—possibly with expert guidance—to better capture lesion structure and improve class separability.

Preprocessing Enhancements: Using lesion localization and contrast enhancement may help the model focus more directly on relevant lesion areas while reducing background noise.

Fine-Tuning Embeddings with Contrastive Learning: using contrastive learning, to better capture lesion-specific representations and improve separability across classes.

Model Ensembling, which may enhance robustness and per-class recall, especially for rare lesion categories.

Real-World Validation, to evaluate model performance on out-of-distribution data and assess generalizability beyond the HAM10000 dataset.

Thank You!