

Automated Skin Lesion Classification - Final Report

Course: DATASCI 281 – Foundations of Computer Vision

Group Members: David Ding, Perry Gabriel, Marcos Geraldo.

Date: August 8th, 2025

1. Proposal

1.1. Problem Statement

Skin cancer is a life-threatening condition where early detection significantly improves prognosis. However, dermatological expertise is often inaccessible in low-resource settings. This motivates the development of automated tools that are fast, interpretable, reliable, and deployable.

1.2. Dataset Characteristics

For this project, we used the **ISIC 2019 Challenge dataset**, which contains **25,331 images** for training and **6,091 images** for testing, across **8 diagnostic categories**.

- **NV** (Melanocytic nevus)
- **MEL** (Melanoma)
- **BCC** (Basal cell carcinoma)
- **BKL** (Benign keratosis)
- **AK** (Actinic keratosis)
- **SCC** (Squamous cell carcinoma)
- **VASC** (Vascular lesion)
- **DF** (Dermatofibroma)

Dataset Link: <https://challenge2019.isic-archive.com>

The characteristics and variations across the images are described below:

- **Resolution:** Images vary from 450×450 to 6000×4000 pixels
- **Channels:** RGB (3 channels)
- **Compression:** JPEG
- **Variation:** Lighting, skin tone, presence of vignettes, artifacts (hairs, rulers)

- **Original Class Distribution:**

Class Name	Training Dataset	Testing Dataset
NV	12876	2495
MEL	4522	1327
BCC	3323	875
BKL	2624	660
AK	867	374
SCC	628	165
VASC	253	104
DF	239	91

- **Final Class distribution:** The final distribution used was **1,000 images per class for training**, making a total of 8,000 images.

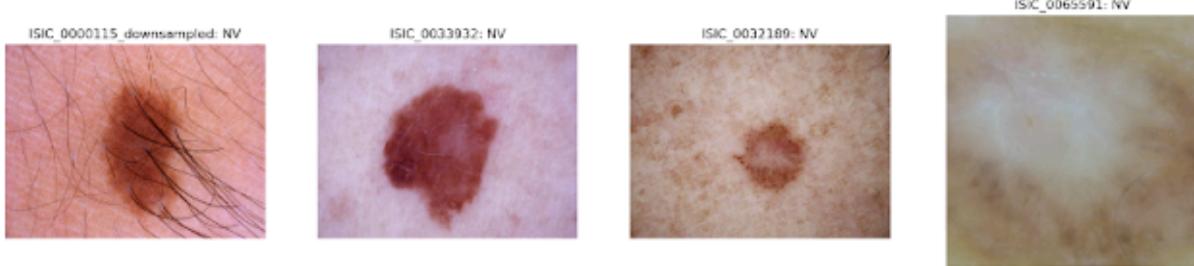
1.3. Intended Classification Problem

We propose a skin lesion classification system able to differentiate skin lesions and to classify them into these **eight categories**, trained on the public medical datasets described above. The model uses both simple image features (e.g., color histograms, HOG) and complex feature embeddings from pre-trained deep CNNs (EfficientNet-B3). We explored multiple combinations of features and models to get the best multi-class classifier performance within the parameters of this project. According to the requirements for this project, we tested a tree-based classifier (**XGBoost**) and a Support Vector Machine (**SVM**).

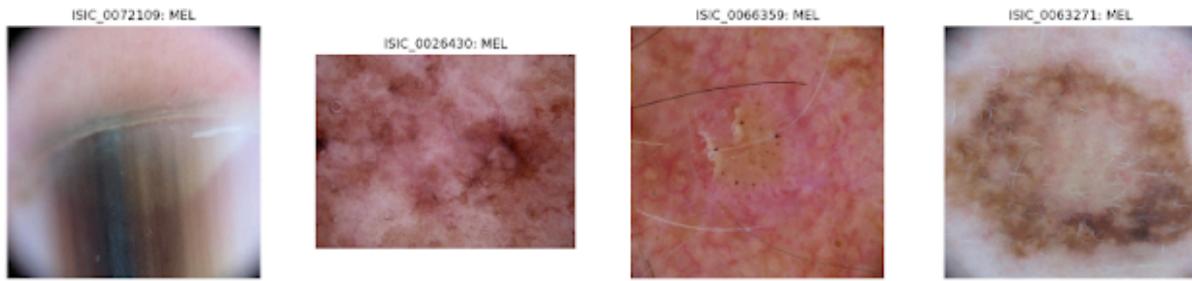
1.4. Image Examples

Below we show a random sample of 4 images for each class from the original training dataset.

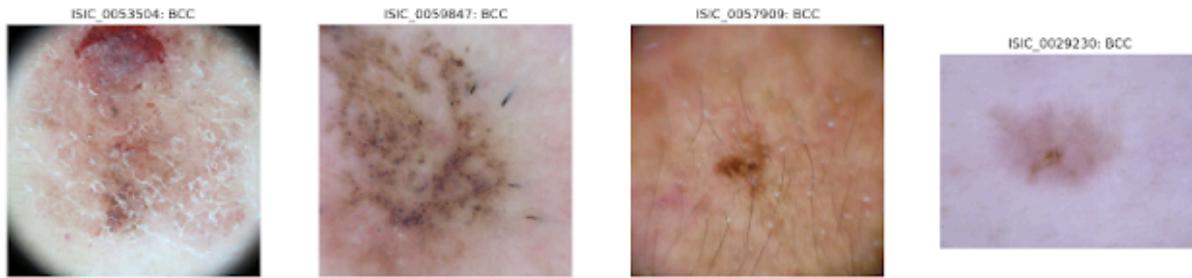
NV (Melanocytic Nevus):



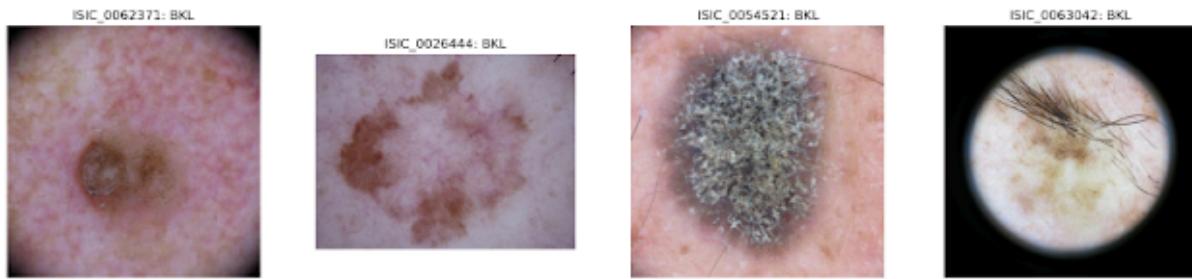
MEL (Melanoma):



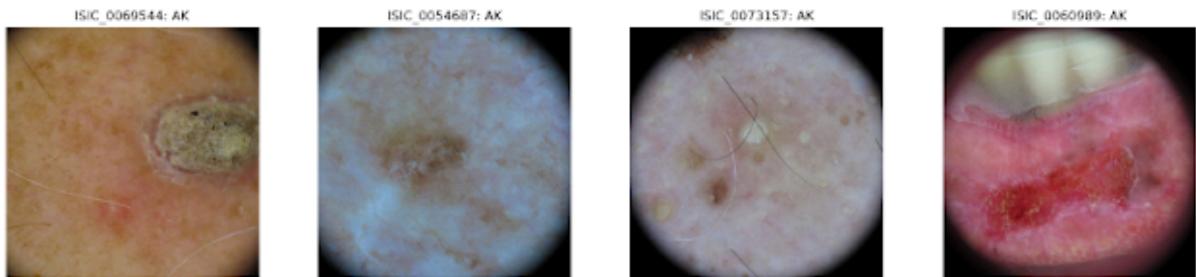
BCC (Basal Cell Carcinoma):



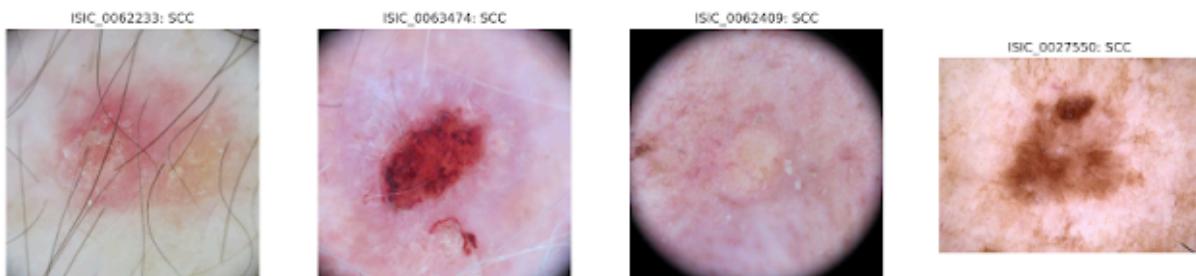
BKL (Benign Keratosis)



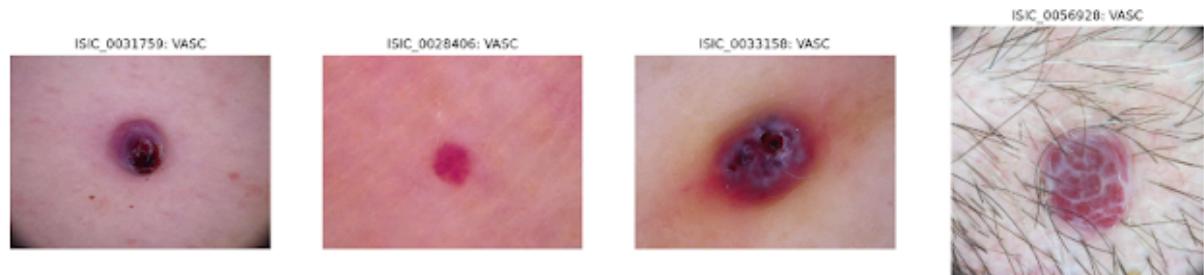
AK (Actinic Keratosis)



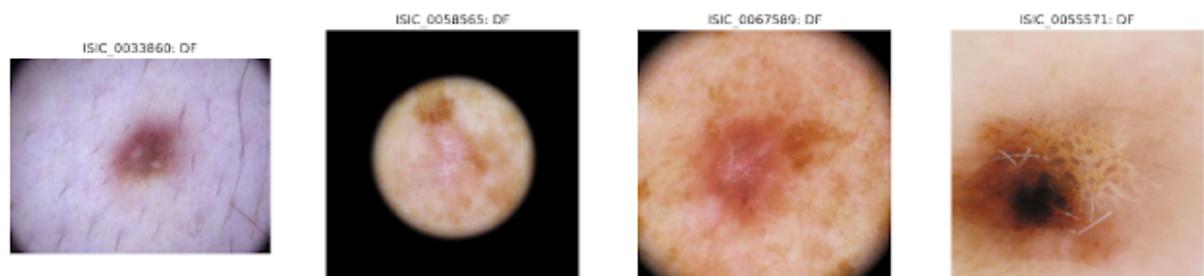
SCC (Squamous Cell Carcinoma)



VASC (Vascular Lesion)



DF (Dermatofibroma)



1.5. Hypothesized Useful Features

Based on preliminary research on medical studies, we have identified specific patterns found in the different types of lesions.

Feature	Melanoma	Carcinoma	Vascular Lesion	Melanocytic nevus (Benign Mole)
Asymmetry	✓ High	✓ Moderate (irregular in SCC)	✗ Rare	✗ Absent
Border Irregularity	✓ High	✓ Possible	✗ Usually smooth	✗ Absent
Color Variation	✓ Significant	✗ Minimal	✓ Distinct (red/purple)	✗ Minimal
Texture Variation	✓ Significant	✓ Possible	✗ Homogeneous	✗ Absent
Specific Dermoscopic Signs	✓ (Pigment network, veil, streaks)	✓ (Telangiectasia, keratin)	✓ (Lacunae, vessels)	✓ (Regular patterns)
Growth Pattern	✓ Rapid	✓ Variable	✓ Can be fast	✗ Stable

Based on these characteristics, we extracted features that would be helpful to differentiate images based on shape, color patterns, and texture.

- **Edges & Gradients:** HOG, Laplacian filters for lesion borders.
- **Texture:** Local Binary Pattern (LBP), GLCM for surface irregularities.
- **Color:** HSV color histograms and metrics for lesion pigmentation.
- **Multi-scale Analysis:** Wavelet decomposition for frequency-based texture.

2. Feature Extraction

2.1. Preprocessing Pipeline

Before feature extraction, we performed a series of steps to standardize the images and to prepare them for processing. This is a summary of those steps:

- **Class Balancing:** We addressed the class imbalance in the training data using the following methods:
 - Random sampling of 1,000 images from each class with more than 1000 elements.
 - Data augmentation by rotation using angles of 90°, 180°, and 270°, and horizontal and vertical image flip.The final distribution used was **1,000 images per class for training.**
- **Image Standardization:** Aspect ratio correction, which sets all images in a 1:1 ratio.
- **Vignette Detection & Removal:** Radial brightness analysis to detect circular shading and cropping.
- **Resizing:** Standardized to 256×256 while preserving lesion integrity.

2.2. Features Implemented

In this project, we selected five groups of features to cover the aspects of Edges, Texture, Color, and Shape. The fifth group is the output of the before-to-last layer of a pre-trained convolutional neural network (EfficientNet B3). The goal of this set of features is to cover the main characteristics that physicians use to diagnose skin lesions. These characteristics are, broadly speaking, the shape of the lesion, color variation, texture, and patterns. Here is the contribution of each feature:

- **Edges & Gradients:**
 - **Histogram of Oriented Gradients (HOG):** Captures lesion boundary gradients.
 - **Laplacian:** Captures lesion borders.
- **Texture and Patterns:**
 - **Local Binary Pattern (LBP):** Encodes lesion texture.
 - **Gray-level Co-occurrence Matrix (GLCM):** Quantifies texture co-occurrence.
 - **Wavelet Decomposition:** Multi-scale lesion texture representation.
- **Color Features:**
 - **HSV color histograms:** Captures distribution of Hue, Saturation, and Value.
 - **HSV metrics:** Captures key values of Hue, Saturation, and value in the lesion.
 - **HSV contrast:** Captures variations of intensity per channel, and as a total.
- **Shape Features:**
 - **Circularity:** Uses the area and perimeter ratio compared to a circle's ratio.
 - **Eccentricity:** The ratio between the longest and shortest axes of the ellipse.
 - **Convexity:** Measures the deviation from a convex shape polygon.
- **CNN:**

- **EfficientNet-B3 penultimate layer:** Extracts the information used by the classification layer of the CNN classifier and uses it as a learned feature.

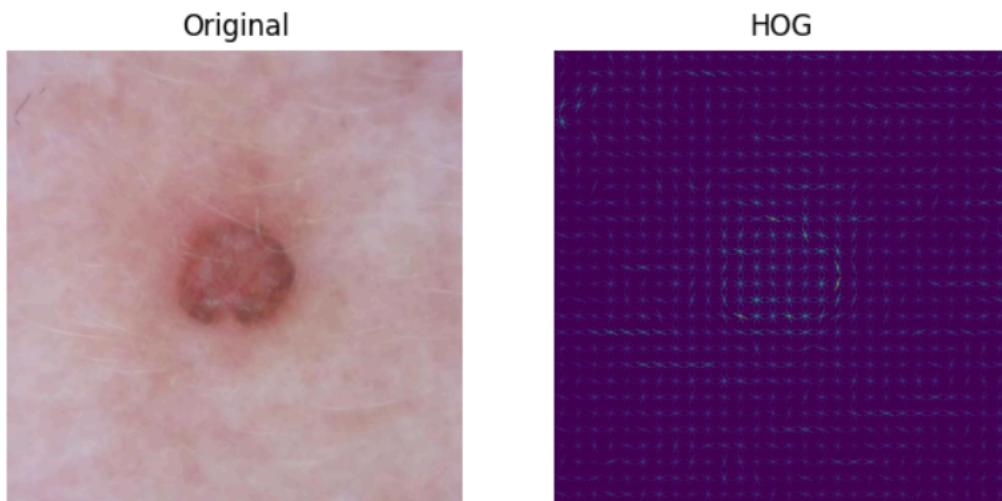
2.3. Visualizations of Features

In this section, we show some examples of each of the five simple features. In the appendix, there is a longer list of examples organized by category for a random selection of images.

HOG (Histogram of Oriented Gradients)

Hog was calculated using the following parameters:

- Orientations = 4
- Pixels per cell = 16 x 16
- Cell per block = 2 x 2
- Block Normalization Method = 'L2-Hys'

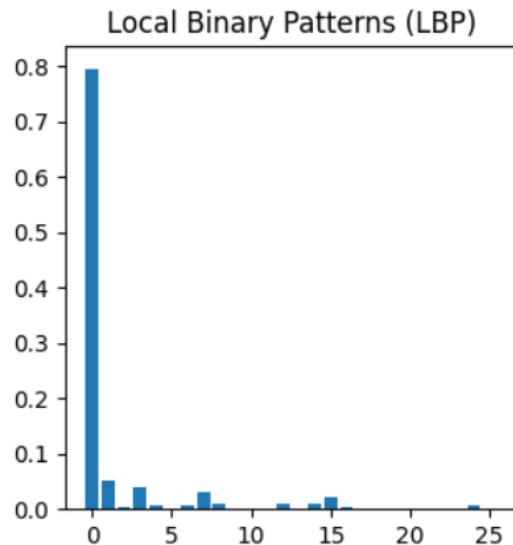


LBP (Local Binary Patterns)

LBP was calculated using the following parameters:

- Radius = 3
- Points = 24

The graph below shows the histogram of binary patterns for the same image above.

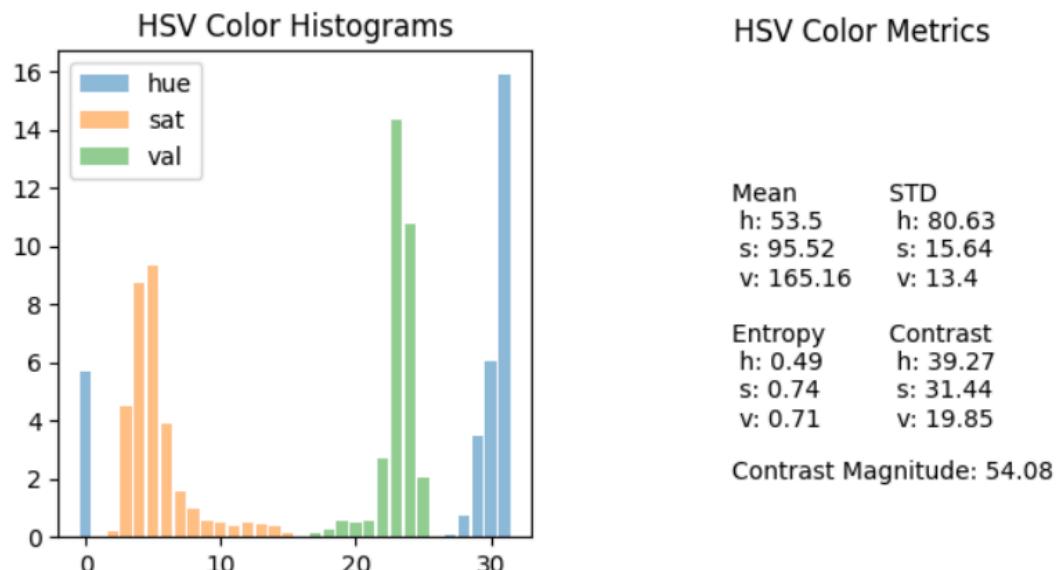


HSV features

We noted that hue and saturation show differences in pigmentation that can help the classification process. To capture those, we used HSV metrics for the image and within the lesion.

The **HSV Color Histogram** was calculated with 32 bins, separately for each channel. The HSV Color Metrics were done by channel, using standard formulae for **mean**, **standard deviation**, and **entropy**.

We also calculated the **contrast per channel** and the **magnitude** of that contrast vector.



GLCM (Gray-Level Co-occurrence Matrix)

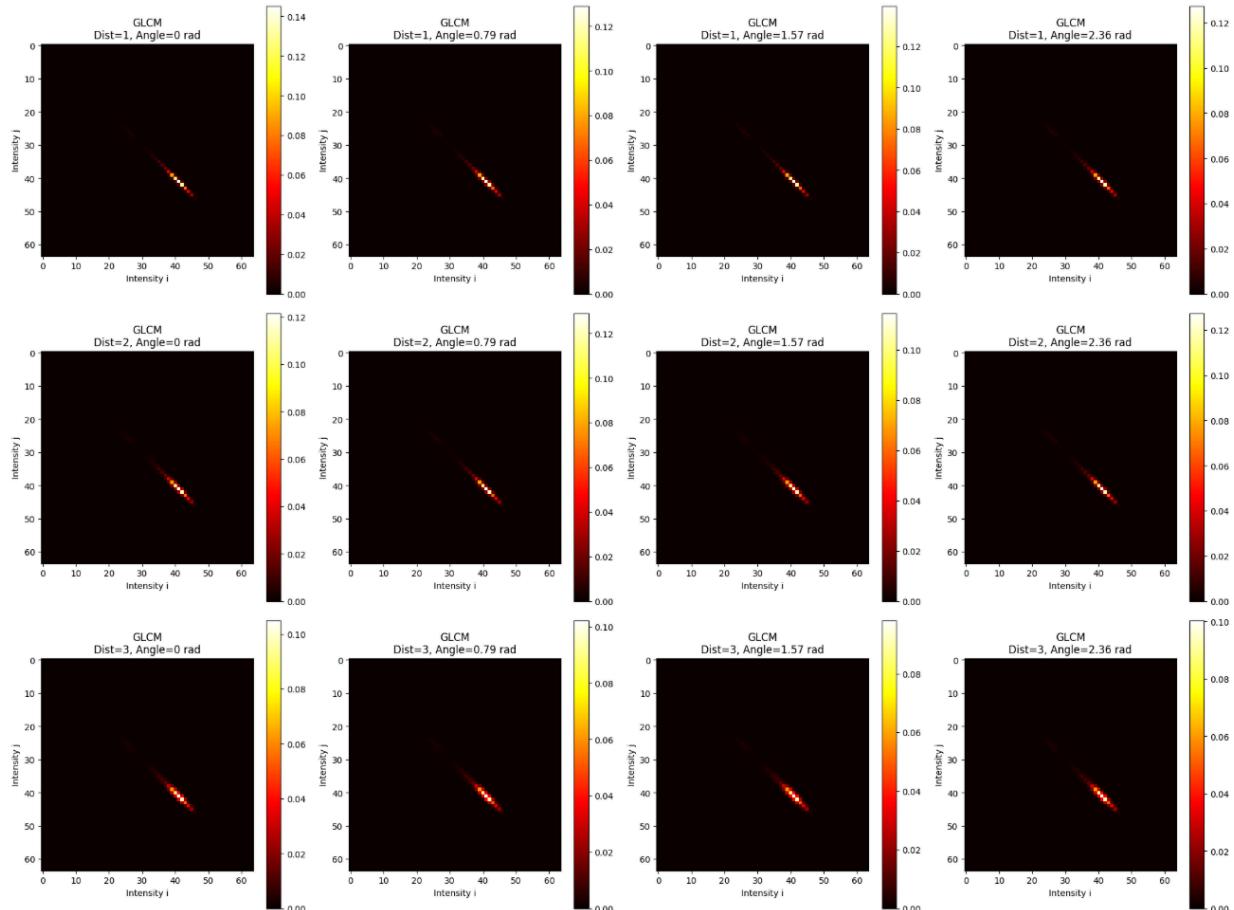
The GLCM features were extracted using the graycomatrix from the skimage package, using the following parameters:

- Distances = 1, 2, and 3.
- Angles = 0° , 45° , 90° , and 135°

For each image we extracted the following metrics:

- Properties = contrast, dissimilarity, homogeneity, energy, correlation, and ASM.

Here is an example of the co-occurrence matrices for the same image as above.

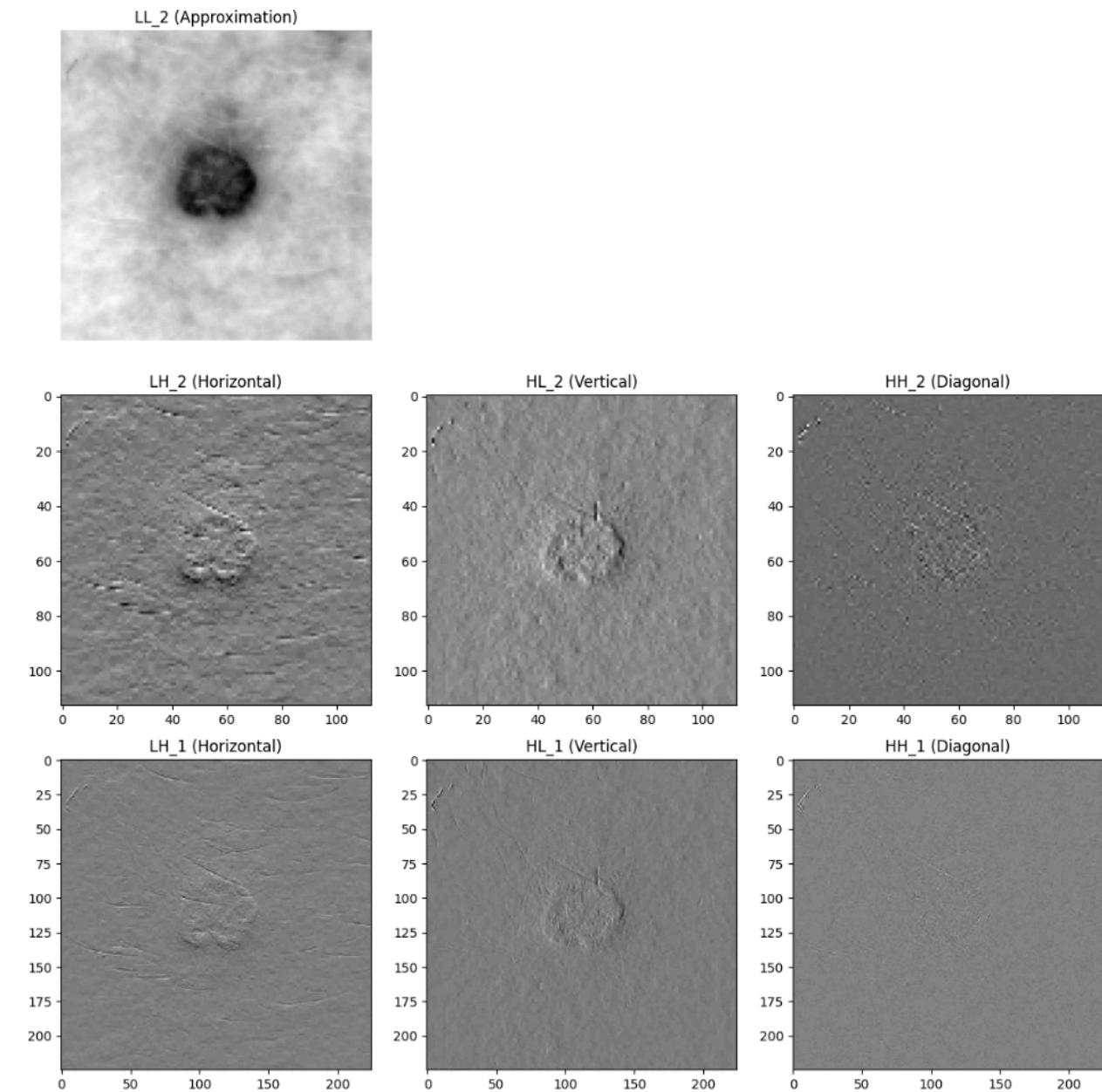


And this is an example of the metrics for the first matrix:

contrast	dissimilarity	homogeneity	energy	correlation	ASM
0.232685	0.211581	0.896208	0.277081	0.993266	0.076774

Wavelet Coefficients

The Wavelet method helps to analyze the image in several frequencies, showing different levels of textures. Here we used two levels, and the respective LL, HL, LH, and HH filters.



Shape Metrics

The detection of the shape of the lesion was done with the function `findContours` and the function `GaussianBlur` from the `cv2` package. The threshold used to determine the contour was the OTSU threshold.

Once the contour of the lesion is found, we can get metrics as area and perimeter, which in turn, allow us to calculate circularity. By fitting an ellipse, we can get the eccentricity. Then, using `convexHull`, we can get the convex polygon that covers the lesion, and calculate the convexity.

```
circularity = (4 * np.pi * area) / (perimeter ** 2)
eccentricity = np.sqrt(1 - (minor_axis / major_axis) ** 2)
convexity = convex_perimeter / perimeter
```

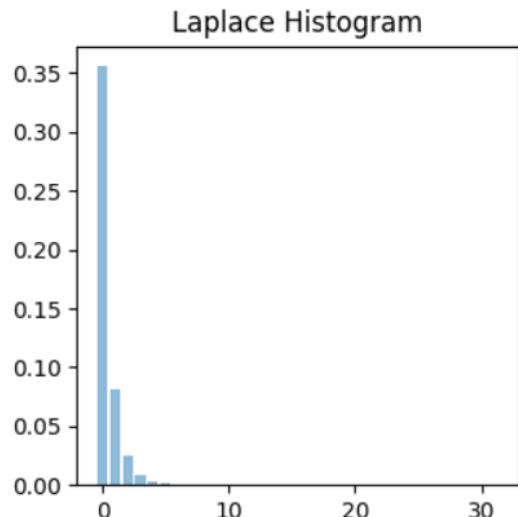
Here are the metrics for the image we have been analyzing.

Shape Metrics

	circularity	eccentricity	convexity
2634	0.426892	0.343492	0.707103

Laplace Histogram

This feature is the histogram of the Laplacian of the image.



CNN features

For the selection of CNN models, we experimented with two different models:

- VGG16
- EfficientNet-B3

We selected EfficientNet-B3 over VGG-16 based on performance.

For the complex features, we used the pre-trained model EfficientNet-B3. The table below shows the general architecture of this model.

Stage	Layer	Input Resolution	Channels	Layers
1	Conv3x3	224 x 224	40	1
2	MBCConv1, k3x3	112 x 112	24	1
3	MBCConv6, k3x3	112 x 112	32	2
4	MBCConv6, k5x5	56 x 56	48	2
5	MBCConv6, k3x3	28 x 28	96	3
6	MBCConv6, k5x5	14 x 14	136	3
7	MBCConv6, k5x5	14 x 14	232	4
8	MBCConv6, k3x3	7 x 7	384	1
9	Conv1x1 + Pooling + FC	7 x 7	1536	1

FINE-TUNING

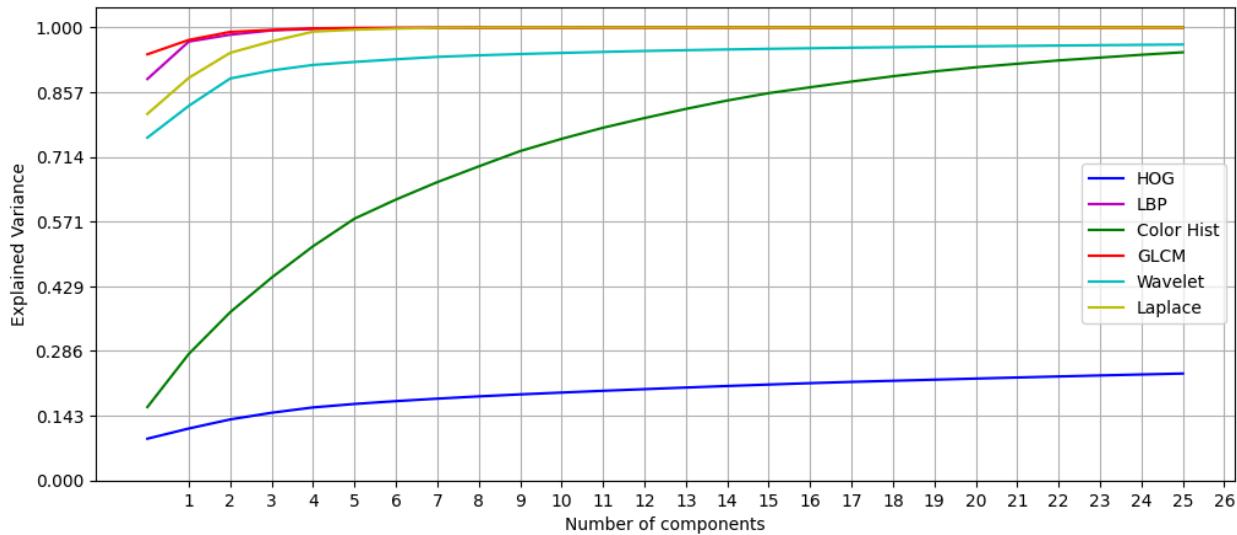
The skin lesion classification model uses a complex progressive training strategy using the EfficientNet-B3 as the backbone architecture. We implemented a two-stage approach which begins with **Stage 1: Frozen Backbone Training**, where only the final classifier layer is trained for 20 epochs while all the remaining backbone parameters remain frozen. This first phase allows the model to adapt to the pre-trained classification model head to the skin lesion domain. This is done while preserving the rich feature representations that were learned from ImageNet. Sidenote, the EfficientNet-B3 model was originally trained on the ImageNet dataset. This training approach uses AdamW optimization with full learning rate, early stopping (patience was set to 10), and class-weighted cross-entropy loss to handle the severe class imbalance present in our dataset.

Stage 2: Full Fine-Tuning unfreezes the entire network and implements differential learning rates to find the most optimal parameters while maintaining stability and performance. This is done by optimizing the classification head to recognize specific skin lesion patterns like asymmetry and color variation. The training process incorporates gradient clipping for stability, cosine annealing scheduling, and medical-focused evaluation that prioritizes cancer detection sensitivity over general accuracy. At this stage, we continuously check for false negative rates and ensure reliable diagnostic performance across all skin lesion types for clinical decision support.

This strategy allows to leverage powerful pre-training features while still adapting them to the specific characteristics of skin lesion imagery.

2.4. Dataset Variation Analysis

- **PCA Decomposition:** For each of the high-dimensional features, we performed a PCA analysis to determine the optimum number of components. The graph below shows a summary of the variability curves of those features:



For each feature, we identified the optimal number of components by two methods.

- Cumulative Variance: The number of components is the one that reaches a cumulative variance of 95% of the total variance.
- Cross Validation: The number of components that returns the best cross-validation score when fitting a logistic regression to the data.

Feature	Components		Graph
	Original	Final	
CNN	1,536	49	<p>Cumulative Explained Variance (First 100 Components)</p> <p>Cumulative Explained Variance</p> <p>Number of Components</p> <ul style="list-style-type: none"> 95.0% threshold Optimal: 49 components
LBP	26	26	<p>Cumulative Explained Variance (First 100 Components)</p> <p>Cumulative Explained Variance</p> <p>Number of Components</p> <ul style="list-style-type: none"> 95.0% threshold Optimal: 26 components

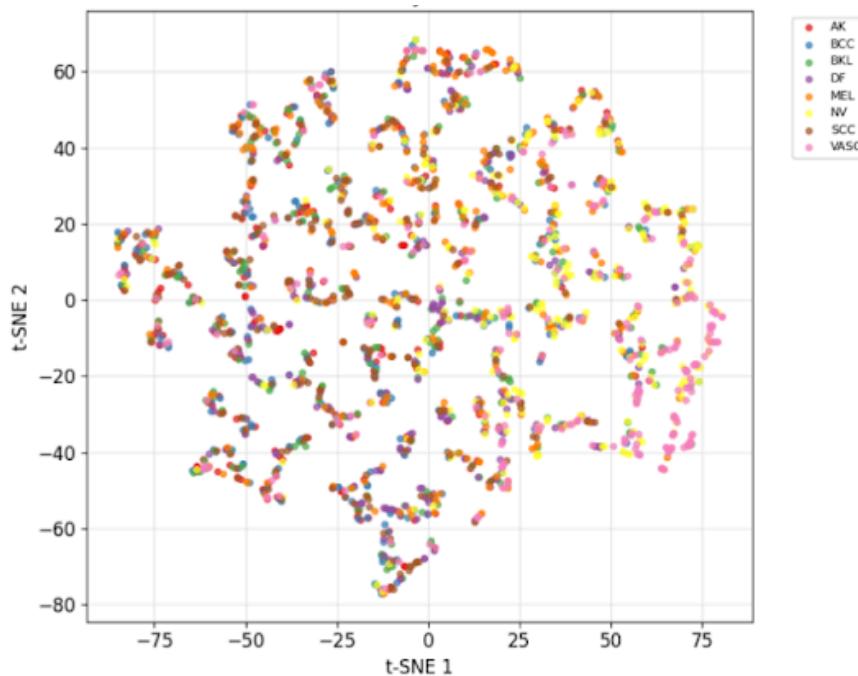
Feature	Components		Graph
	Original	Final	
Wavelet	1,000	14	<p>Cumulative Explained Variance (First 100 Components)</p> <p>95.0% threshold</p> <p>Optimal: 14 components</p>
Laplace Hist	32	32	<p>Cumulative Explained Variance (First 100 Components)</p> <p>95.0% threshold</p> <p>Optimal: 32 components</p>

Feature	Components		Graph
	Original	Final	
GLCM	72	49	<p>Cumulative Explained Variance (First 100 Components)</p> <p>Cumulative Explained Variance</p> <p>Number of Components</p> <p>95.0% threshold</p> <p>Optimal: 49 components</p>
HOG	11,664	17	<p>Cumulative Explained Variance (First 100 Components)</p> <p>Cumulative Explained Variance</p> <p>Number of Components</p> <p>95.0% threshold</p> <p>Optimal: 17 components</p>

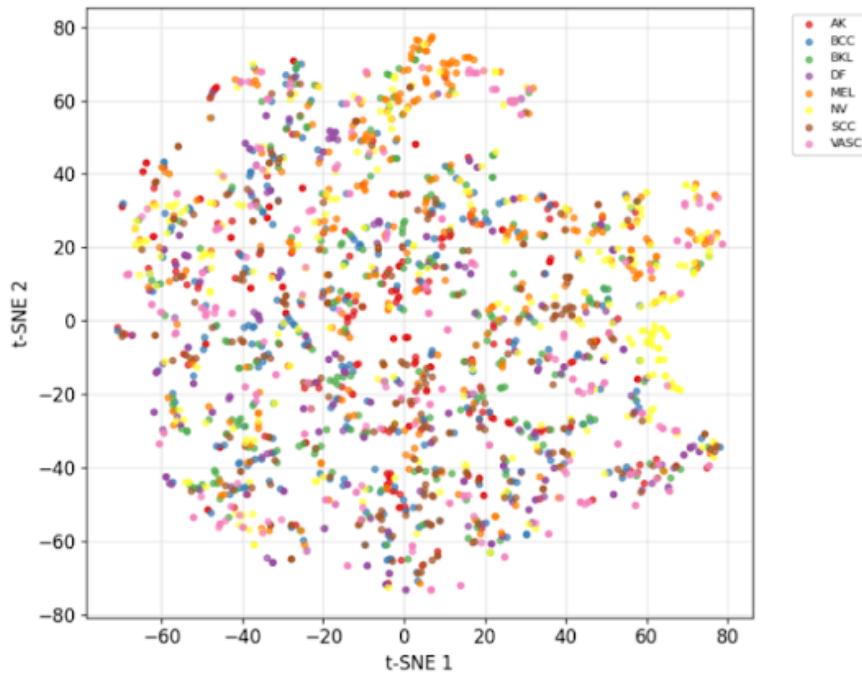
Feature	Components		Graph
	Original	Final	
HSV	96	41	

- t-SNE Visualization

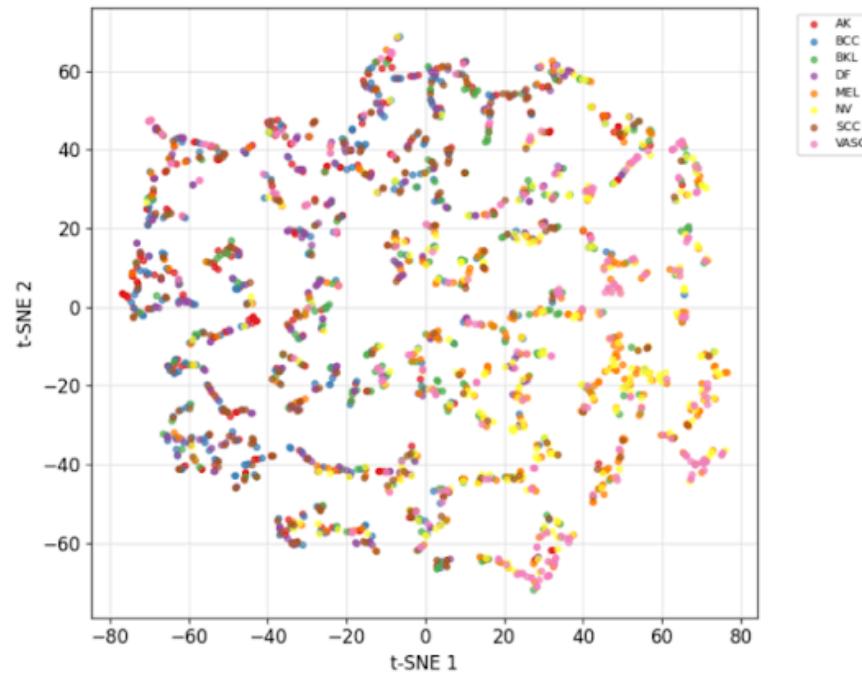
Circular Features: The graph of t-SNE projection for circular features shows some weak alignment of points by class, with the strongest one being the VASC class.



Color Features: The color features show some weak alignment for the NV class, and heavy overlap on the other classes.



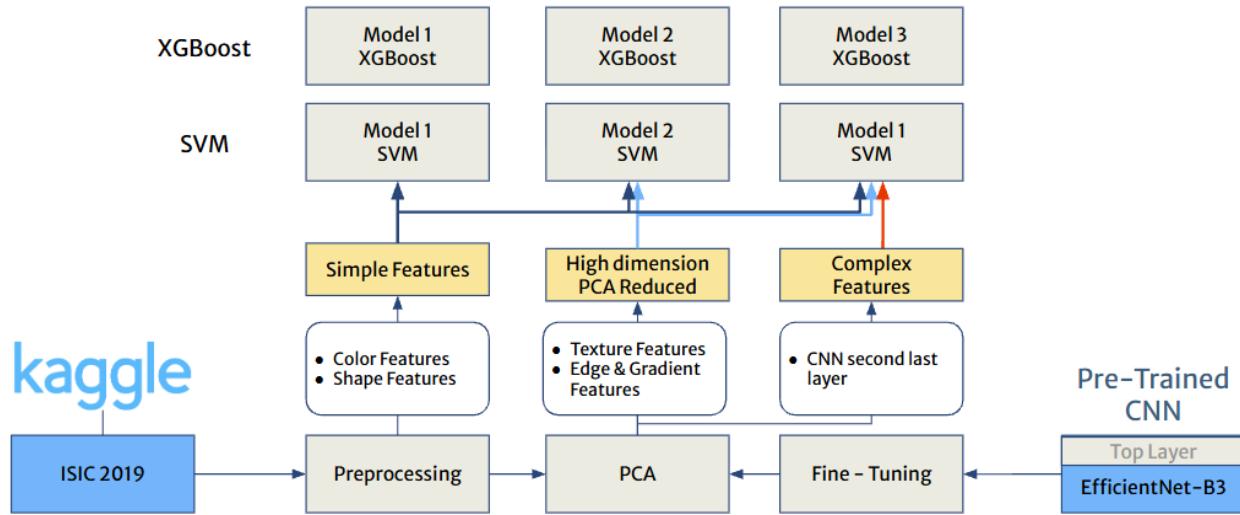
Contrast Features: The contrast features show some alignment for the VASC abd NV classes.



3. Classification

3.1. Approach

The classification was done using two classification algorithms (XGBoost and SVM), and three sets of features (Simple, High Dimensional, Complex), making a total of 6 models. The diagram below shows the feature groups and how they are distributed in each of the 6 models.



Where

- **Simple Features:** Are metrics of low-dimensional features that don't require PCA reduction and are used directly in the model.
 - Color Features:
 - HSV color histograms
 - HSV metrics
 - HSV contrast
 - Shape Features:
 - Circularity
 - Eccentricity
 - Convexity
- **High-Dimensional Features:** Are high-dimensional features that went through the process of PCA reduction described above.
 - Edges & Gradients:
 - HOG
 - Laplacian
 - Texture and Patterns:
 - Local Binary Pattern (LBP)
 - GLCM

- Wavelet decomposition
- **Complex Features:** The output of the before-last layer of the EfficientNet-B3 pre-trained model.
 - CNN features

3.2. Classifiers & Methodology

XGBoost and SVM selection rationale:

In medical classification, ensemble methods mimic methods used by professionals, adding subtle feature combinations within each sample to determine and distinguish between benign and malignant lesions. XGBoost's methods follow a similar train of thought, capturing and combining multiple weak decision points to create and capture complex feature interactions.

Support Vector Machines are also well-suited for high-dimensional classification problems. There are likely complex decision boundaries that medical specialists will undergo within their diagnostic process; therefore, SVM's resilience to dimensionality should better capture that level of separation.

Dataset Decomposition:

The dataset was systematically divided into a training and testing set with an 80/20 ratio using stratified resampling to ensure a balanced representation of all eight skin lesion classes in both the training and testing sets. The stratification was crucial for two reasons: the inherent class imbalance in typical dermatological datasets, and due to the data augmentation procedures which had already created synthetic samples in the dataset. The training set consisted of 6400 samples, and the testing set contained 1600 samples, maintaining the original class distribution ratios.

The following section will now cover model-specific cross-validation approaches and results.

XGBoost

Through cross-validation, the hyperparameters for XGBoost were optimized for the three models individually through 3-fold cross-validation. The XGBoost hyperparameter search space included `n_estimators` [10, 200, 300], `max_depth` [4, 6, 8], learning rate [0.05, 0.1, 0.2], and `subsample` [0.8, 0.9, 1.0], resulting in 81 parameter combinations per model. The grid search revealed that XGBoost models consistently favored models with moderate complexity, almost always using `max_depth` values of 6-8.

The following model summary shows the performance of the three XGBoost models.

DETAILED XGB MODEL COMPARISON

Model	Features	Best CV Score	Test Accuracy	F1	Test ROC-AUC	Training Time
Model 1	16	0.6359	0.6819	0.6721	0.9254	31.18s
Model 2	195	0.6331	0.6900	0.6828	0.9324	317.27s
Model 3	244	0.6556	0.6931	0.6859	0.9357	403.23s

OPTIMAL HYPERPARAMETERS

Model 1 (Non-PCA only):

```
learning_rate: 0.2
max_depth: 8
n_estimators: 300
```

Model 2 (Non-PCA + PCA w/o EfficientNet):

```
learning_rate: 0.2
max_depth: 8
n_estimators: 300
```

Model 3 (All features):

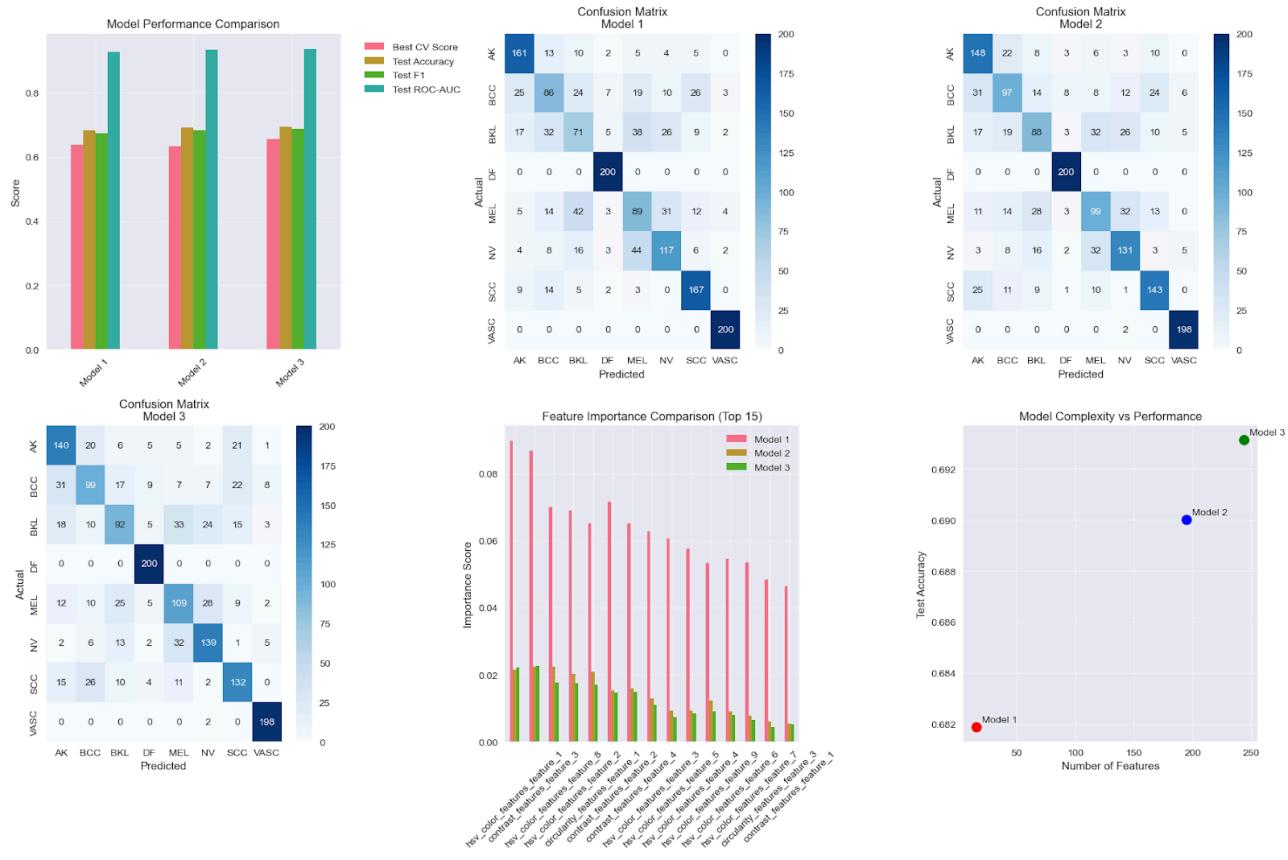
```
learning_rate: 0.2
max_depth: 6
n_estimators: 300
```

Best performing model: Model 3 (All features)

Test accuracy: 0.6931

Best CV score: 0.6556

The next image shows a few diagnostic plots for the XGBoost modelling. The final models were able to classify VASC and DF classes incredibly well, with increasing performance going from simpler to more complex models. There were fewer misclassifications between MEL, NV, SCC classes as the feature complexity increased.



Support Vector Machines (SVMs)

Again, through cross-validation, several hyperparameter configurations were experimented with in the SVM modelling process. The grid search explored kernel types ['rbf', 'poly', 'sigmoid'], regularization parameters [0.1, 1, 10, 100], and gamma values ['scale', 'auto', 0.001, 0.01, 0.1, 1]. This totalled to 72 combinations per model. After tuning, the SVM models predominantly selected RBF kernels with varying regularization and gamma parameters. This highlights the algorithm's internal sensitivities to feature scaling and regularization strength.

DETAILED SVM MODEL COMPARISON

	Model	Features	Best CV Score	Test Accuracy	F1 Test	ROC-AUC	Training Time
Model 1		16	0.6252	0.6863	0.6841	0.9267	180.50s
Model 2		195	0.5469	0.5775	0.5763	0.8929	437.13s
Model 3		244	0.6142	0.6562	0.6572	0.9034	530.08s

OPTIMAL HYPERPARAMETERS

Model 1 (Non-PCA only):

C: 10
gamma: 1
kernel: rbf

Model 2 (Non-PCA + PCA w/o EfficientNet):

C: 10
gamma: scale
kernel: rbf

Model 3 (All features):

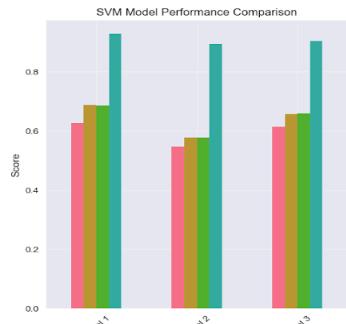
C: 100
gamma: scale
kernel: poly

Best performing SVM model: Model 1 (Non-PCA only)

Test accuracy: 0.6863

Best CV score: 0.6252

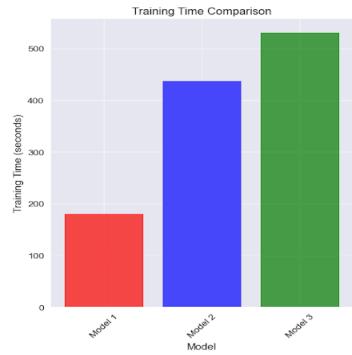
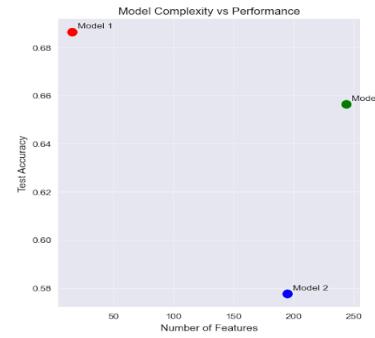
Again the following images detail the SVM performance. Uniquely with the SVM models, we see Model 1 to be the clear winner in terms of model performance. The additional features did not contribute to gains in performance.



Confusion Matrix Model 1									
		Predicted							
		AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Actual	AK	155	12	10	1	12	4	6	0
	BCC	26	78	31	2	36	10	15	0
	BKL	13	25	66	4	63	25	4	0
	DF	0	1	0	199	0	0	0	0
	MEL	6	9	29	1	126	22	7	0
	NV	2	7	12	0	65	112	2	0
	SCC	3	9	7	0	10	4	167	0
	VASC	0	0	0	0	5	0	0	195

Confusion Matrix Model 2									
		Predicted							
		AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Actual	AK	109	30	18	9	9	1	22	2
	BCC	37	82	26	13	9	6	19	8
	BKL	19	16	82	3	32	32	13	3
	DF	3	9	6	171	1	2	8	0
	MEL	16	15	35	8	89	25	8	4
	NV	3	12	20	3	32	124	0	6
	SCC	38	27	8	8	21	3	94	1
	VASC	4	3	7	0	5	7	1	173

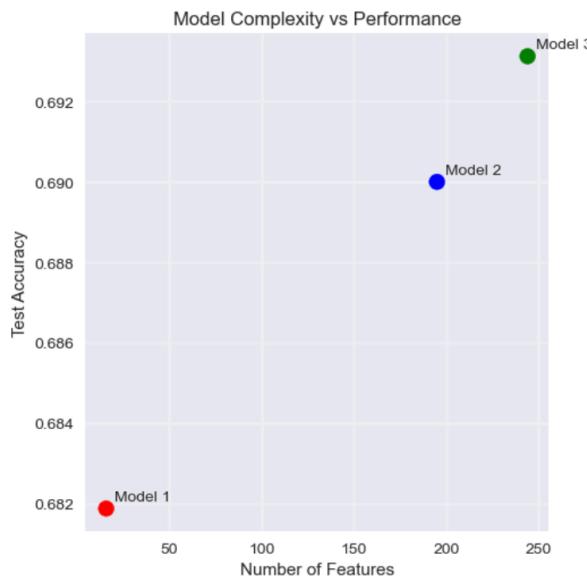
Confusion Matrix Model 3									
		Predicted							
		AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Actual	AK	117	35	12	0	6	0	29	1
	BCC	19	130	13	7	7	4	20	0
	BKL	27	16	95	1	21	17	22	1
	DF	2	6	1	183	4	1	3	0
	MEL	12	25	27	6	99	17	10	4
	NV	3	12	22	2	29	123	1	8
	SCC	18	30	7	4	7	3	129	2
	VASC	1	10	6	0	4	2	3	174



4. Efficiency - Accuracy Trade-off

The performance of the XGBoost versions of the models was superior to each of their SVM counterparts, so we will focus the discussion on the XGBoost models.

Among the three XGBoost models, the selection of the best model depends on the selection criteria. The graph below maps each model based on complexity, measured as the number of dimensions in the features, and accuracy.



It is important to note that not all the dimensions cost the same. For example, the additional dimensions coming from the Complex feature (CNN) not only require more computing power to obtain, but also require additional preprocessing and fine-tuning of the pre-trained model.

In the same way, execution time would miss all the additional preprocessing and fine-tuning needed to start using the CNN model. In this case, Model 3 also includes all the features in Models 1 and 2, so this metric is an acceptable proxy to measure relative cost, with the caveat that the gap between

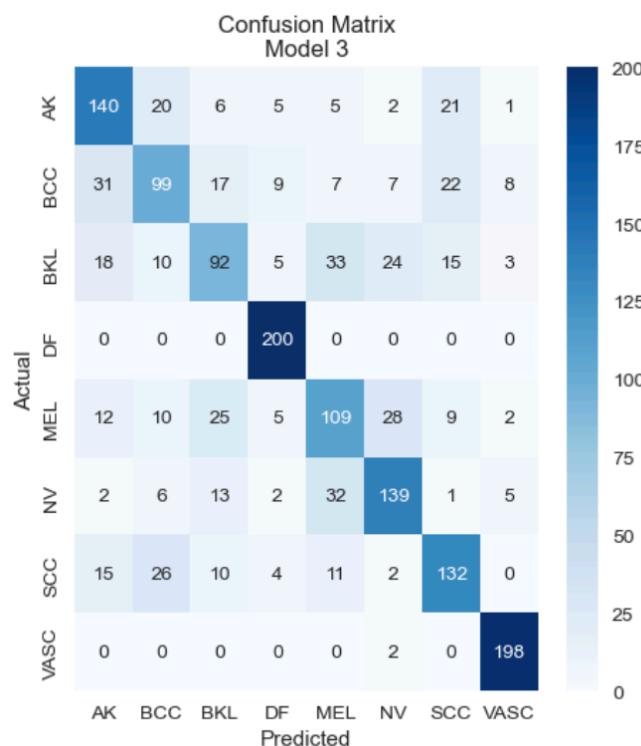
Model 3 and the rest is actually higher.

	Best Performance	Best Efficiency
Model name	Model 3 XGBoost	Model 1 XGBoost
Number of Features	244	16
Accuracy	0.6931	0.6819
F1	0.6875	0.6721
ROC-AUC	0.9357	0.9254
Training Time	403.23s	31.18s
Features included	Color Features Shape Features Texture & Pattern Features Edges & Gradient Features CNN	Color Features Shape Features

Best Performance: Model 3 XGBoost

The table below shows the performance of the best performance model by each class, and the corresponding confusion matrix.

XGBoost Model 3	Class	Recall	Precision	Accuracy	F1
Actinic Keratosis	AK	0.7000	0.6422	0.9138	0.6699
Basal Cell Carcinoma	BCC	0.4950	0.5789	0.8919	0.5337
Benign Keratosis	BKL	0.4600	0.5644	0.8881	0.5069
Dermatofibroma	DF	1.0000	0.8696	0.9813	0.9302
Melanoma	MEL	0.5450	0.5533	0.8881	0.5491
Melanocytic Nevus	NV	0.6950	0.6814	0.9213	0.6881
Squamous Cell Carcinoma	SCC	0.6600	0.6600	0.9150	0.6600
Vascular Lesion	VASC	0.9900	0.9124	0.9869	0.9496



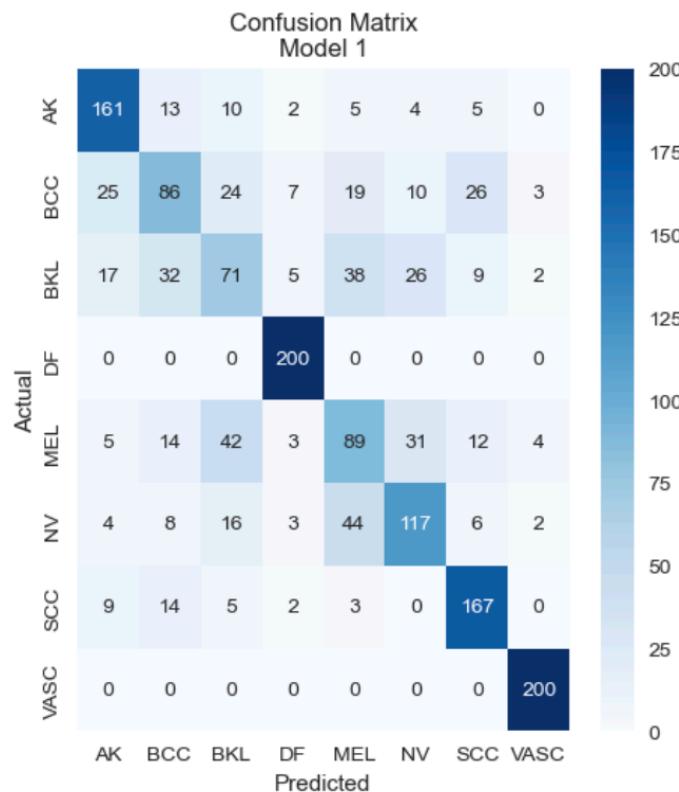
This model is very strong in identifying Vascular Lesions (VASC) and Dermatofibroma (DF) with recall of 99% and 100% respectively.

The biggest confusion happens between BKL (Benigne Keratoma) and Melanoma (MEL). These two kinds of lesions share color and shape characteristics, with cases where they mimic each other.

We also see classification errors among BCC (Basal Cell Carcinoma), SCC (Squamous Cell Carcinoma), and AK (Acnitic Keratosis). The confusion is expected since Acnitic Keratosis is a precursor of Squamous Carcinoma and shares some characteristics with it.

Best Efficiency: Model 1 XGBoost

XGBoost Model 1		Class	Recall	Precision	Accuracy	F1
Actinic Keratosis		AK	0.8050	0.7285	0.9380	0.7648
Basal Cell Carcinoma		BCC	0.4300	0.5244	0.8798	0.4725
Benign Keratosis		BKL	0.3550	0.4226	0.8585	0.3859
Dermatofibroma		DF	1.0000	0.9009	0.9862	0.9479
Melanoma		MEL	0.4450	0.4495	0.8622	0.4472
Melanocytic Nevus		NV	0.5939	0.6223	0.9054	0.6078
Squamous Cell Carcinoma		SCC	0.8350	0.7422	0.9430	0.7859
Vascular Lesion		VASC	1.0000	0.9479	0.9931	0.9732



Similarly as Model 3, Model 1 is very strong in identifying Vascular Lesions (VASC) and Dermatofibroma (DF) with recall of 100% for both categories.

The biggest confusion in this case happens between Melanoma (MEL) and Melanocytic Nevus (NV).

We also see errors between Benign Keratosis (BKL) and Melanoma with a high number of cases of melanoma misclassified as benign.

There are smaller classification errors among BCC (Basal Cell Carcinoma), SCC (Squamous Cell Carcinoma), and AK (Acnitic Keratosis).

5. Discussion

The main challenge for these models is to learn multiple patterns and shapes of skin lesions, which might be the reason behind the classification errors seen in the confusion matrices. Even though we have 8 classes, each of them has its own heterogeneity.

For example, the confusion between BK and MEL might be caused because within the class **Benign Keratosis (BKL)**, there are Seborrheic Keratosis, Keratoacanthoma, and Keratosis Pilaris (Follicular Keratosis). Each of these has different characteristics that the model needs to learn.

It is also common that benign keratosis lesions share characteristics with lesions from other keratosis that are considered cancer precursors like Acnitic Keratosis (AK). This model makes a good job differentiating those two classes

On the other hand, there are cases when Keratomas can mimic Melanomas and vice versa, sharing similar color patterns and irregular shapes that produce the confusion. This might be the reason for the relatively high confusion between MEL and BKL.

We have addressed this problem as a classification exercise, but in medical implementations we would recommend adjusting the thresholds for the classification to optimize by Precision or Recall depending on the context of the exercise and whether it is important to false positives or false negatives.

Another challenge was the presence of artifacts such as hair and rulers in the images. We experimented with DullRazor, a method to remove hair from the images that proved to be useful to detect contours, but damaged textures, so we couldn't apply it to all features.

6. Conclusion & Future Work

Our study shows that XGBoost models performed comparably across increasing dimensionality, and outperformed SVM models.

The best-performing model, XGBoost Model 3, which included CNN and traditional features, achieved strong class-level performance, particularly for *Vascular Lesions* (VASC) and *Dermatofibromas* (DF), with recall rates of 100% in test. However, confusion persisted between visually and structurally similar classes such as *Benign Keratosis* (BKL) and *Melanoma* (MEL), and among *Basal Cell Carcinoma* (BCC), *Squamous Cell Carcinoma* (SCC), and *Actinic Keratosis* (AK), reflecting the similarities of characteristics among these lesions.

The study also shows that while complex features improve accuracy, simpler models with fewer features offer competitive results at a fraction of the computational cost, making them attractive for practical deployment scenarios where efficiency is critical. The most efficient model, XGBoost Model 1, showed a small decline in accuracy compared to Model 3. The kind and

magnitude of errors though are more concerning (it confuses Melanocytic Nevus (Moles) and Melanomas), requiring a careful threshold optimization to ensure the reduction of False Negatives, at the expense of increasing False Positives.

Overall, this work confirms that hybrid feature strategies leveraging both computer vision and deep learning methods can yield robust performance in dermatological image classification.

Future work may focus on improving the feature extraction to capture the subtle differences between related classes, handling better artifacts, and adjusting the decision thresholds of the classifiers to align the model outputs with diagnostic priorities.

7. References

References:

[1] Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5, 180161 doi:10.1038/sdata.2018.161 (2018)

[2] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)", 2017; arXiv:1710.05006.

[3] Hernández-Pérez C, Combalia M, Podlipnik S, Codella NC, Rotemberg V, Halpern AC, Reiter O, Carrera C, Barreiro A, Helba B, Puig S, Vilaplana V, Malvehy J. BCN20000: Dermoscopic lesions in the wild. *Scientific Data*. 2024 June 17;11(1):641.

[4] Gao, J., Fei, W., Shen, C., Shen, X., Sun, M., Xu, N., Li, Q., Huang, C., Zhang, T., Ko, R., Cui, Y., & Yang, C. (2021). Dermoscopic Features Summarization and Comparison of Four Types of Cutaneous Vascular Anomalies. *Frontiers in Medicine*, 8, 692060.

<https://doi.org/10.3389/fmed.2021.692060>

[5] Ciudad-Blanco C, Avilés-Izquierdo JA, Lázaro-Ochaita P, Suárez-Fernández R. Dermoscopic findings for the early detection of melanoma: an analysis of 200 cases. *Actas Dermosifiliogr*. 2014 Sep;105(7):683-93. English, Spanish. doi: 10.1016/j.ad.2014.01.008. Epub 2014 Apr 3. PMID: 24704190.

<https://pubmed.ncbi.nlm.nih.gov/24704190/>

[6] Lupu M, Caruntu C, Popa MI, Voiculescu VM, Zurac S, and Boda D: Vascular patterns in basal cell carcinoma: Dermoscopic, confocal and histopathological perspectives (Review). *Oncol Lett* 17: 4112-4125, 2019.

<https://www.spandidos-publications.com/10.3892/ol.2019.10070/abstract>

[7] Argenziano G, Zalaudek I, Corona R, et al. Vascular structures in skin tumors: a dermoscopy study of 531 lesions. Arch Dermatol. 2004;140(12):1485–1489.
doi:10.1001/archderm.140.12.1485. PMID: 15611426.
<https://pubmed.ncbi.nlm.nih.gov/15611426/>

8. Appendix

8.1. Lesion's Characteristics

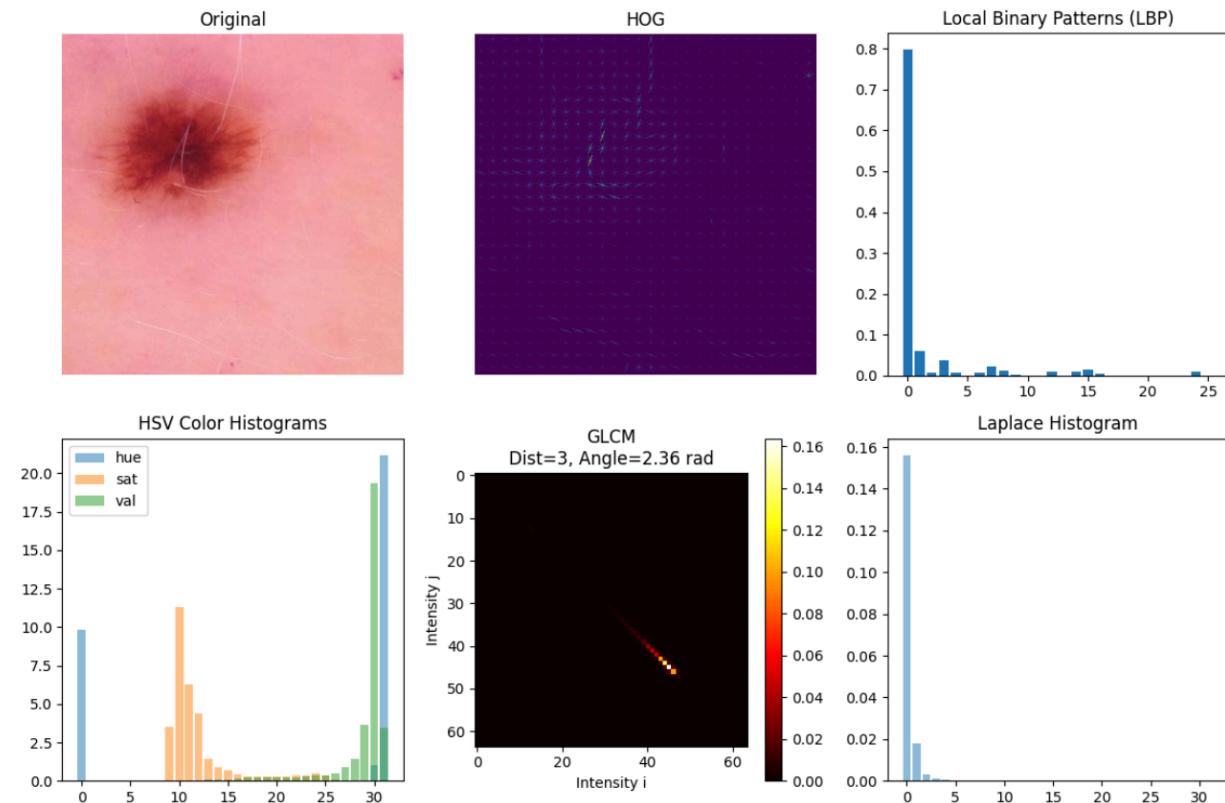
Lesion	Typical Clinical Appearance	Typical Dermoscopic Cues	Reference
Actinic Keratosis (AK)	Rough, scaly erythematous patch or plaque; often on sun-exposed skin; may be pigmented; <1 cm; background photodamage.	Red pseudonetwork; white/yellow scales; targetoid hair follicles; rosettes (polarized); strawberry pattern (erythema + follicular openings with white halos).	Argenziano et al., 2004 ; Ciudad-Blanco et al., 2014
Basal Cell Carcinoma (BCC)	Pearly, translucent papule/nodule with rolled border; may ulcerate; telangiectasias; sometimes pigmented.	Arborizing vessels; blue-gray ovoid nests; leaf-like areas; spoke-wheel areas; ulceration; multiple blue-gray dots; shiny white-red structureless areas.	Lupu et al., 2019 ; Argenziano et al., 2004
Benign Keratosis (BKL)	“Stuck-on” waxy or verrucous plaque; sharply demarcated; various colors (tan to black).	Milia-like cysts; comedo-like openings; fissures/ridges (“brain-like” surface); moth-eaten borders; hairpin vessels; sharp demarcation; fat-finger pattern.	Argenziano et al., 2004 ; Tschanl et al., 2018
Dermatofibroma (DF)	Firm papule or nodule; dimples with lateral pressure; often brownish with central lighter area.	Central white scar-like area; delicate peripheral pigment network; dotted vessels; peripheral light brown pigmentation; crystalline structures.	Argenziano et al., 2004
Melanoma (MEL)	Asymmetric pigmented lesion; irregular borders; color variegation; flat or nodular; ABCD features.	Atypical pigment network; irregular streaks; blue-white veil; regression structures (white scar-like, peppering); irregular dots/globules; multiple colors; atypical vascular patterns.	Ciudad-Blanco et al., 2014
Melanocytic Nevus (NV)	Symmetric pigmented macule or papule; smooth borders; stable over time.	Symmetric pigment network; regular dots/globules; homogeneous or reticular patterns; pseudopods/streaks (regular); uniform color distribution.	Tschanl et al., 2018
Squamous Cell Carcinoma (SCC)	Hyperkeratotic papule/plaque/nodule; may ulcerate; sun-exposed skin; indurated.	Keratin masses; white structureless areas; hairpin or glomerular vessels; surface scale; irregular vessels; central ulcer/crust.	Argenziano et al., 2004
Vascular Lesion (VASC)	Red/violaceous papule, nodule, or plaque; blanchable; may bleed easily.	Dotted vessels; linear irregular vessels; lacunae (red, purple, blue-black); homogeneous red/purple areas; absence of pigment network.	Gao et al., 2021 ; Argenziano et al., 2004

8.2. Image Examples:

This section contains a random example for each class and the visualization of each feature.

NV (Melanocytic nevus)

Image id: 406
Class:NV



HSV Metrics

	Mean	Std	Entropy	Contrast
h	99.509654	87.740493	0.527379	35.284822
s	172.620301	22.568447	0.802925	51.089291
v	174.907398	31.106881	0.854427	48.468502

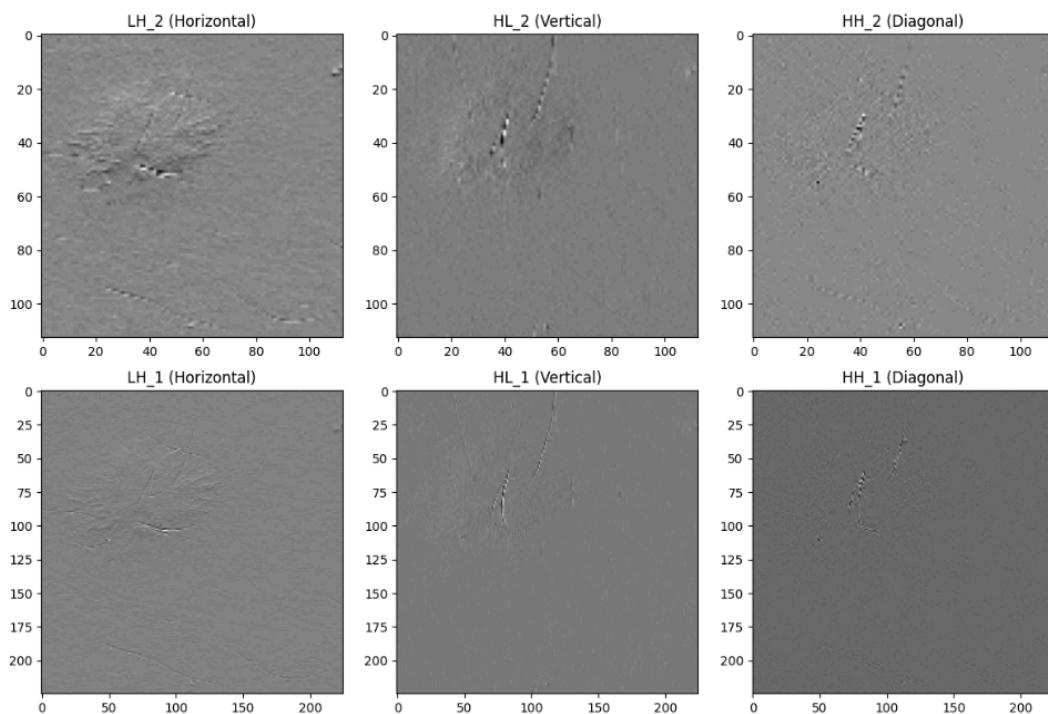
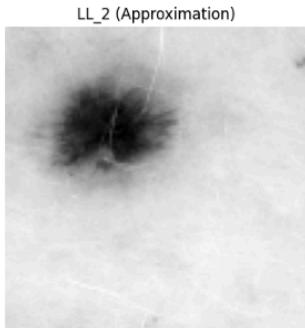
GLCM Metrics

	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.59806	0.312438	0.85993	0.267132	0.99491	0.071359

Shape Metrics

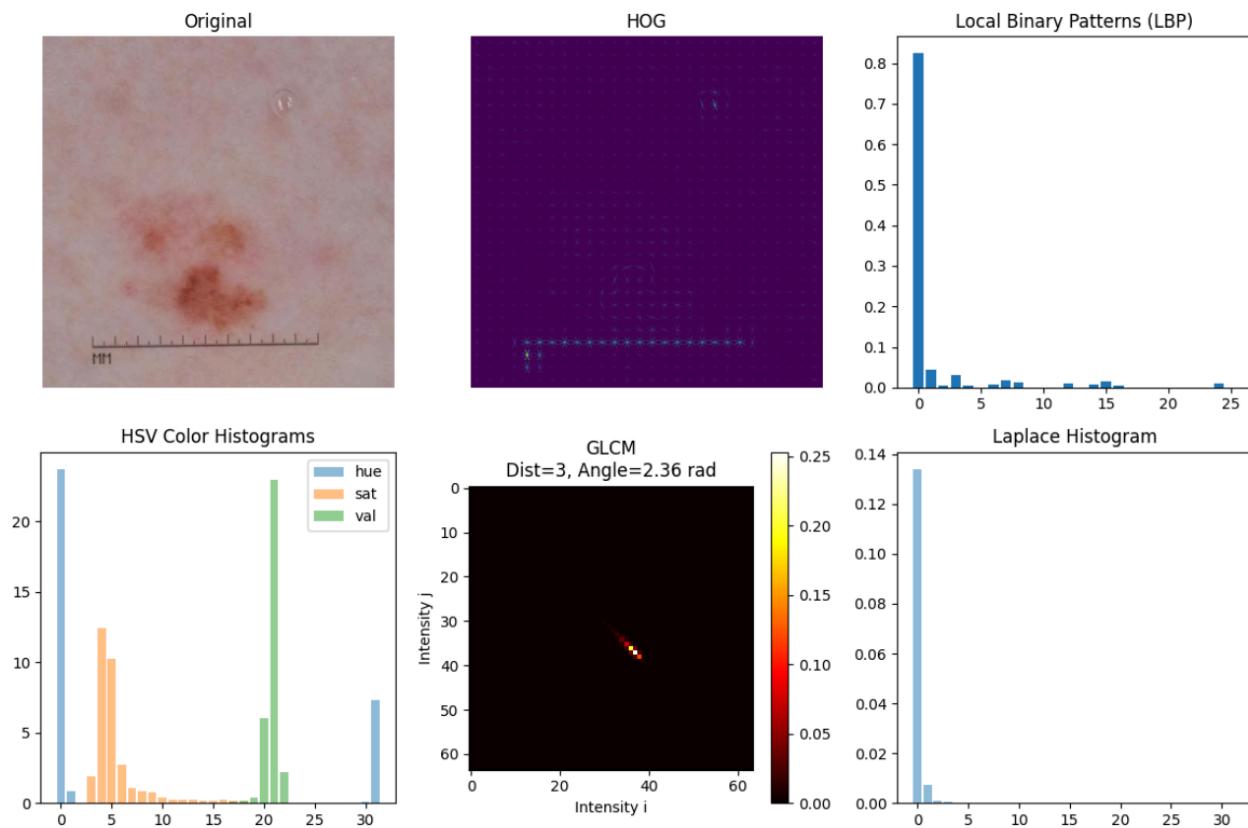
	circularity	eccentricity	convexity
406	0.467033	0.662981	0.740768

Wavelet



MEL (Melanoma)

Image id: 1349
Class:MEL



HSV Metrics

	Mean	Std	Entropy	Contrast
h	4.882150	7.265006	0.399267	11.103166
s	116.406044	20.405761	0.791511	45.640592
v	162.894698	8.570057	0.633952	10.523804

GLCM Metrics

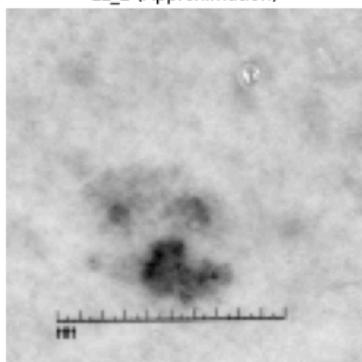
	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	1.085439	0.334457	0.864976	0.360908	0.930103	0.130255

Shape Metrics

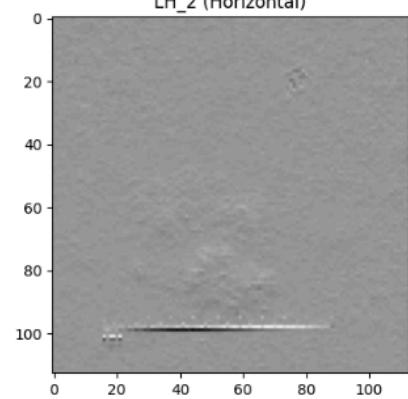
	circularity	eccentricity	convexity
1349	0.238189	0.501667	0.621168

Wavelet

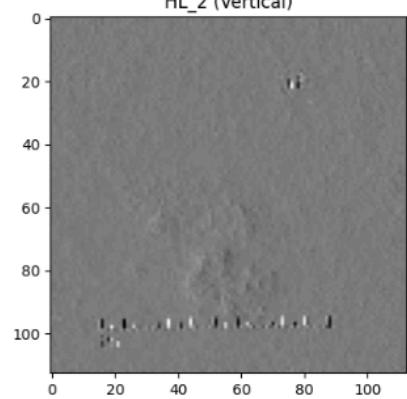
LL_2 (Approximation)



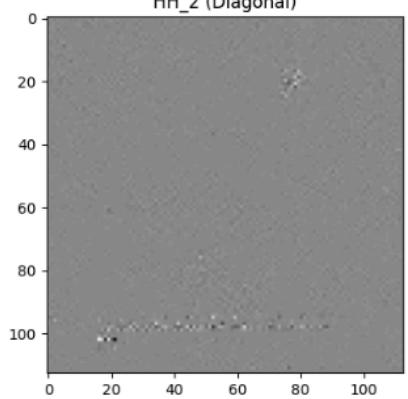
LH_2 (Horizontal)



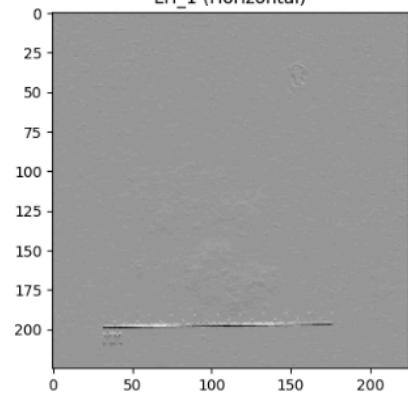
HL_2 (Vertical)



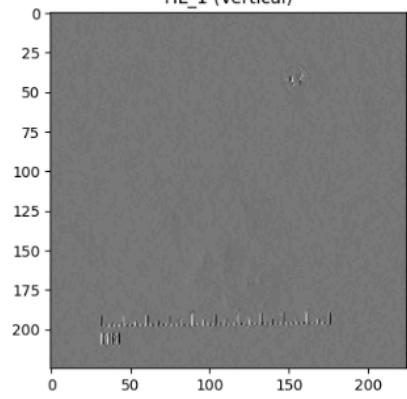
HH_2 (Diagonal)



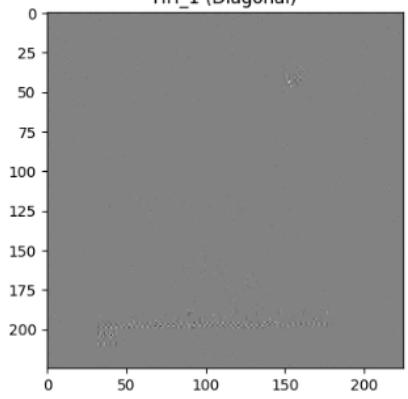
LH_1 (Horizontal)



HL_1 (Vertical)

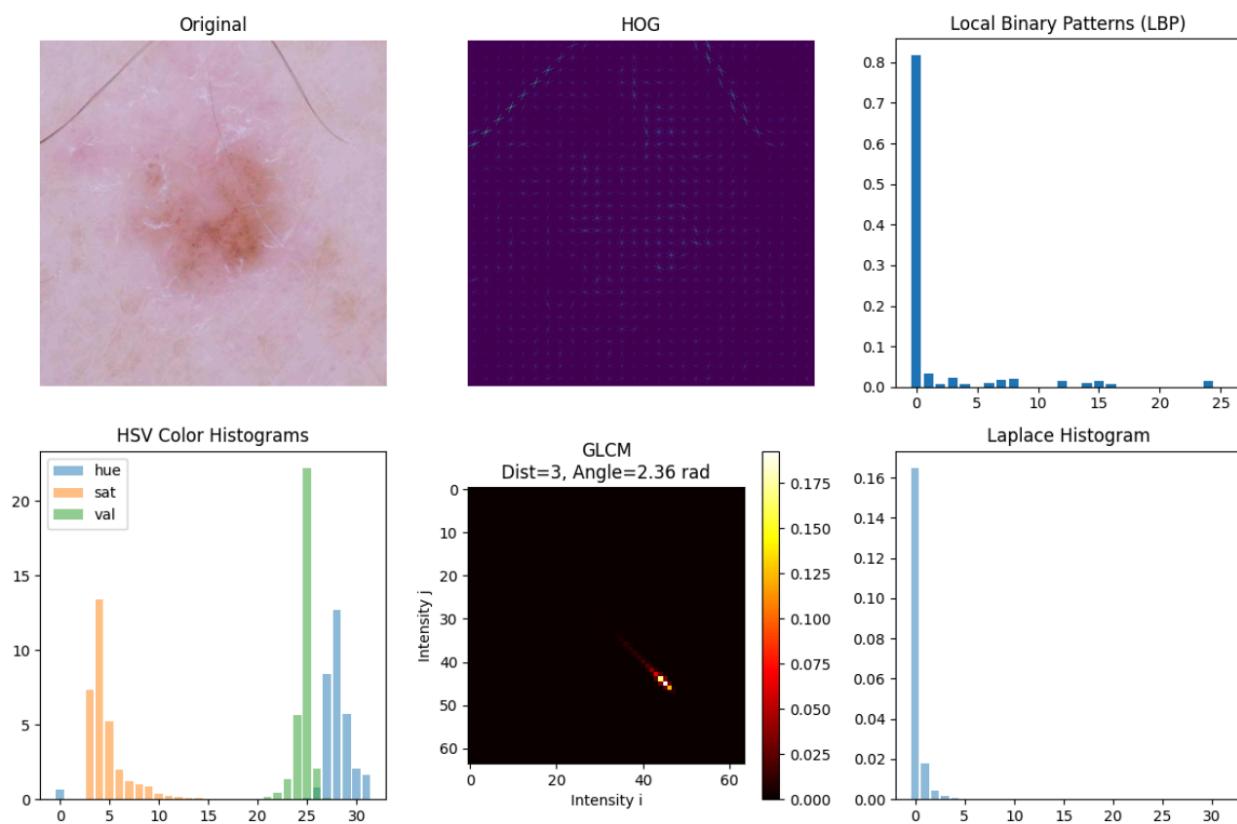


HH_1 (Diagonal)



BCC (Basal cell carcinoma)

Image id: 2082
Class:BCC



HSV Metrics

	Mean	Std	Entropy	Contrast
h	142.315885	65.519928	0.599206	21.037552
s	72.426609	14.747499	0.726840	23.866042
v	191.374188	6.983563	0.605369	8.823388

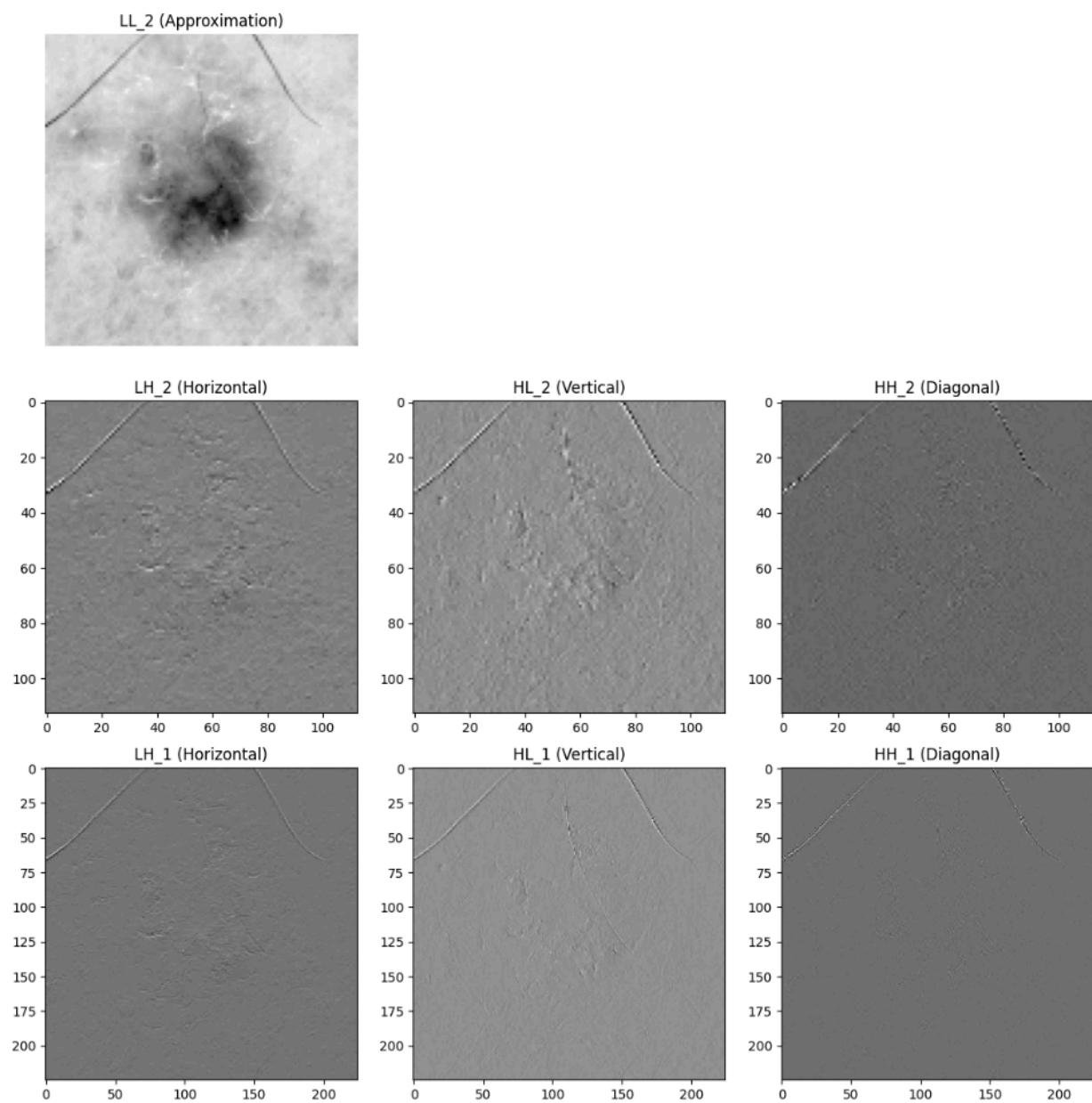
GLCM Metrics

	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.539119	0.350275	0.839482	0.306468	0.973259	0.093922

Shape Metrics

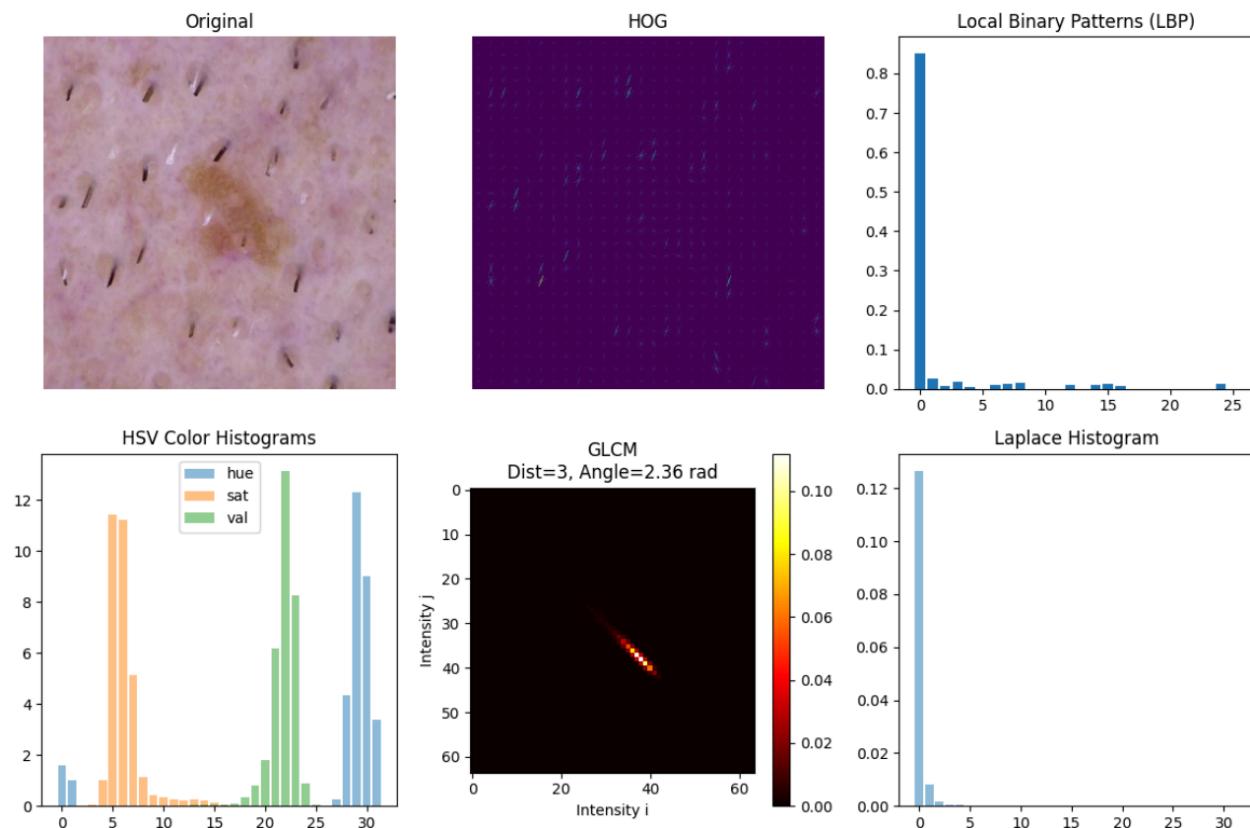
	circularity	eccentricity	convexity
2082	0.27814	0.576856	0.580437

Wavelet



BKL (Benign keratosis)

Image id: 3054
Class:BKL



HSV Metrics

	Mean	Std	Entropy	Contrast
h	33.216569	61.382855	0.588068	112.908431
s	94.647490	19.893889	0.788186	37.475752
v	159.891247	12.450400	0.651483	17.366543

GLCM Metrics

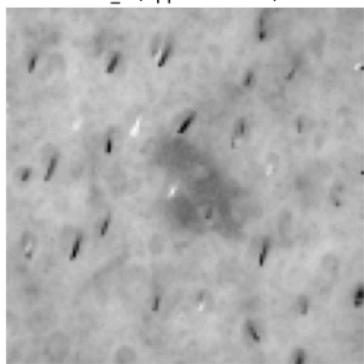
	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	1.286998	0.507221	0.786688	0.232246	0.946469	0.053938

Shape Metrics

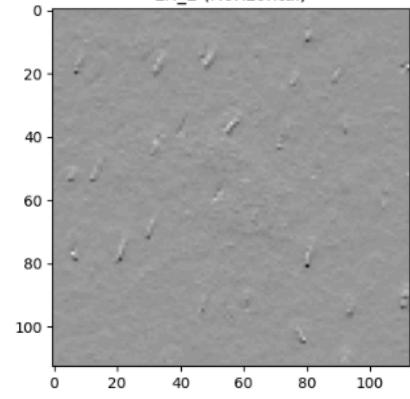
	circularity	eccentricity	convexity
3054	0.225518	0.765891	0.615942

Wavelet

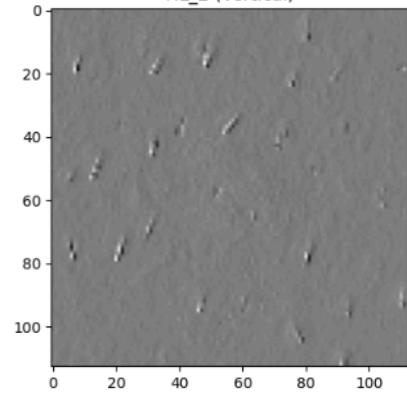
LL_2 (Approximation)



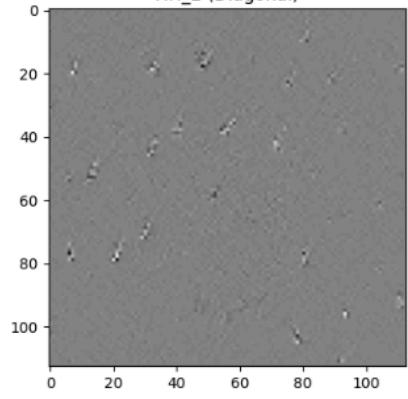
LH_2 (Horizontal)



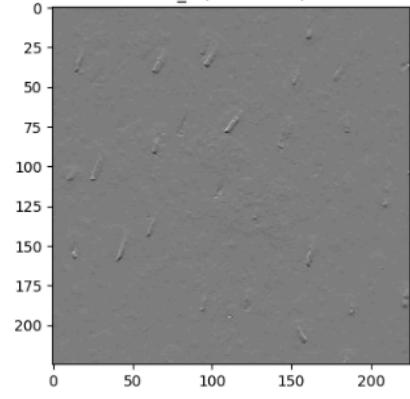
HL_2 (Vertical)



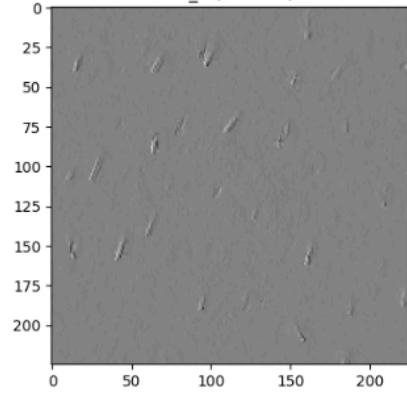
HH_2 (Diagonal)



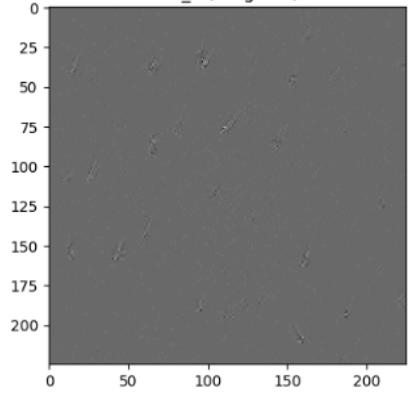
LH_1 (Horizontal)



HL_1 (Vertical)

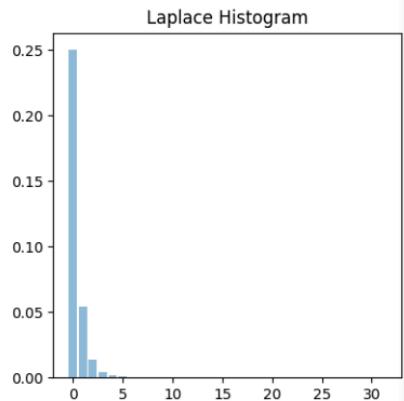
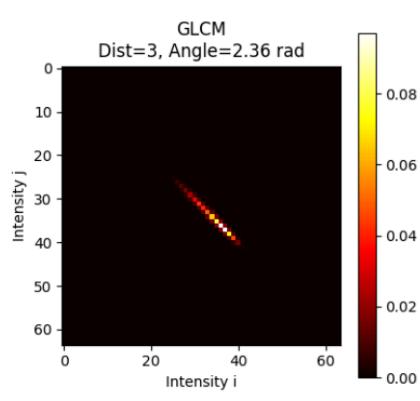
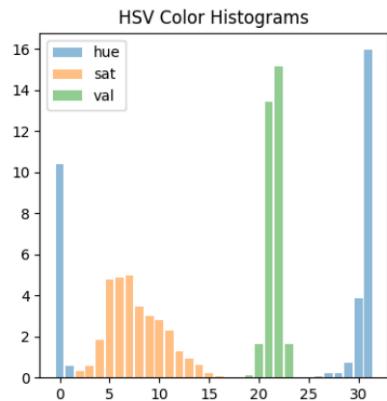
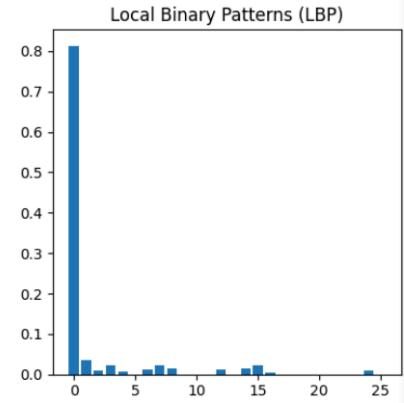
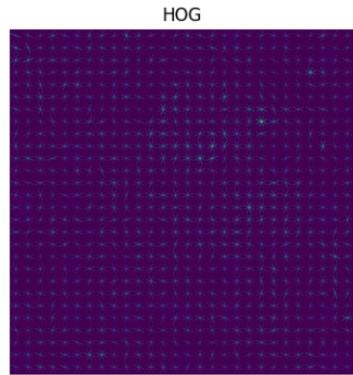


HH_1 (Diagonal)



AK (Actinic keratosis)

Image id: 4967
Class: AK



HSV Metrics

	Mean	Std	Entropy	Contrast
h	101.895163	86.888687	0.540689	19.313699
s	90.284037	13.619335	0.720561	26.331746
v	172.535911	4.644460	0.549851	5.349957

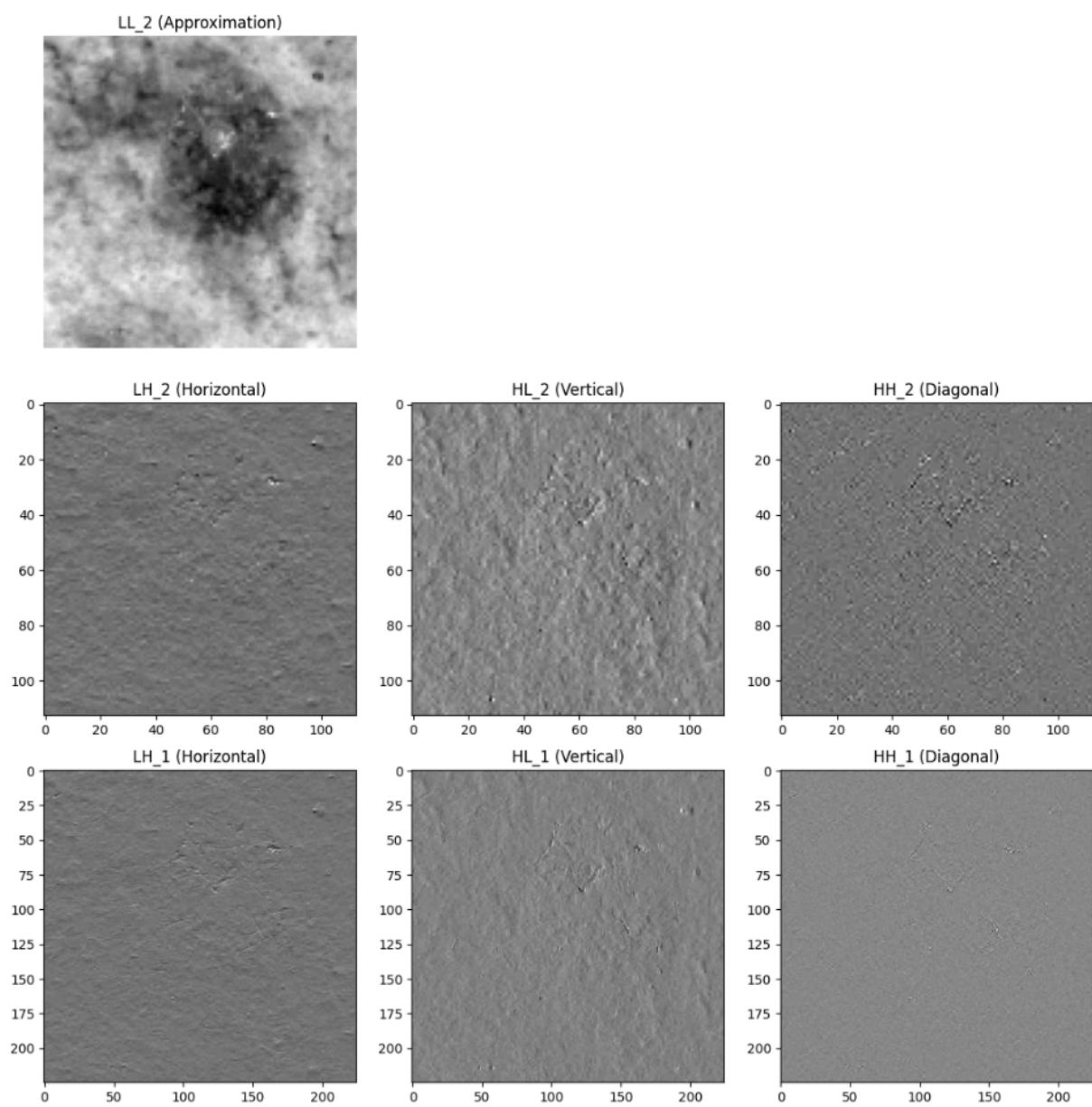
GLCM Metrics

	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.354259	0.31419	0.846687	0.220716	0.984087	0.048716

Shape Metrics

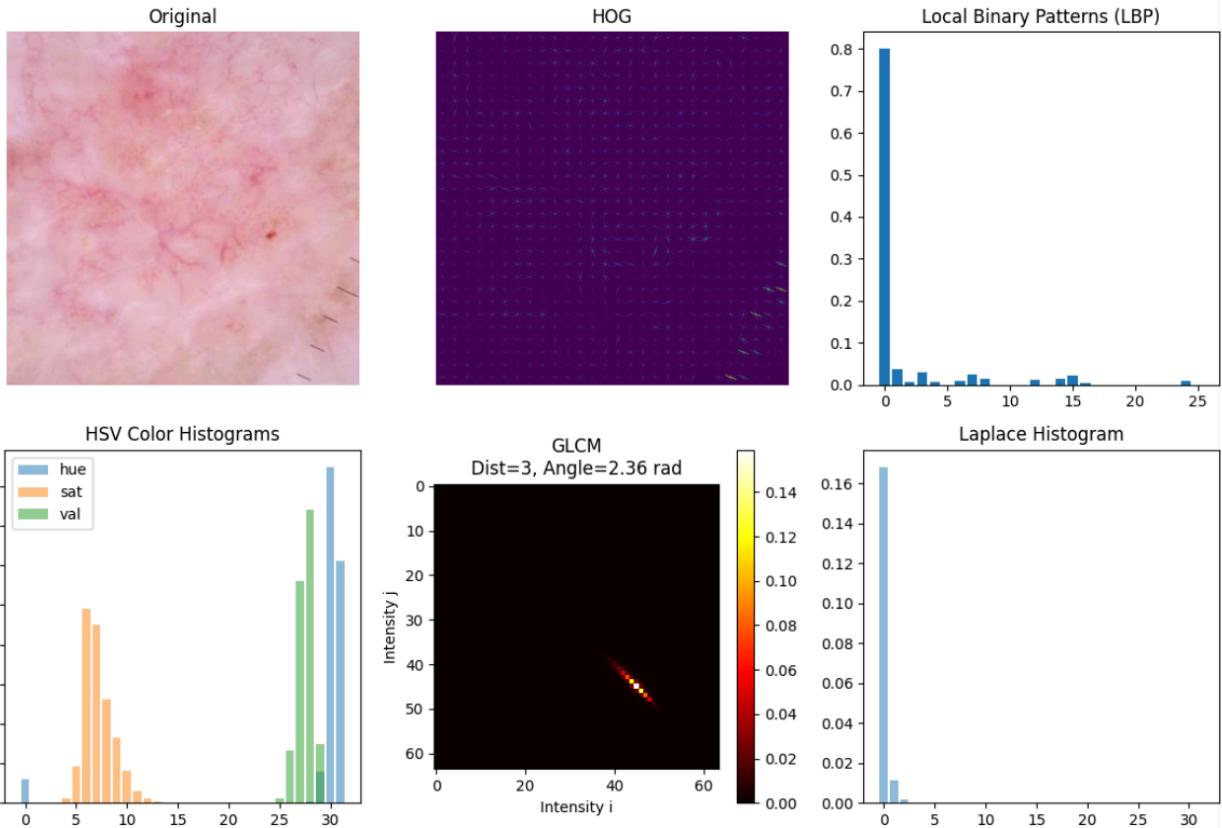
	circularity	eccentricity	convexity
4967	0.118533	0.724626	0.448978

Wavelet



SCC (Squamous cell carcinoma)

Image id: 5532
Class:SCC



HSV Metrics

	Mean	Std	Entropy	Contrast
h	169.473876	28.243665	0.494908	2.031124
s	76.359592	9.852355	0.666108	14.800974
v	223.107085	5.706802	0.577744	4.131676

GLCM Metrics

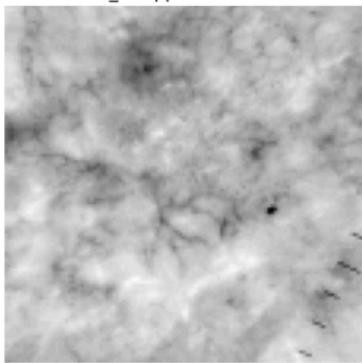
	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.344637	0.286731	0.861396	0.278767	0.970064	0.077711

Shape Metrics

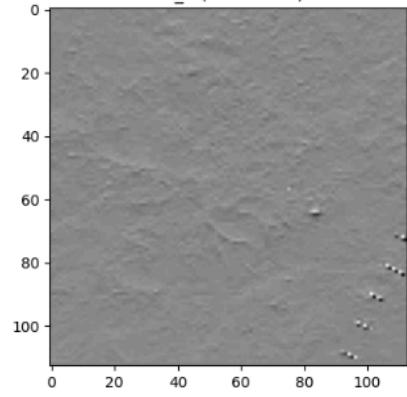
	circularity	eccentricity	convexity
5532	0.033084	0.571951	0.272695

Wavelet

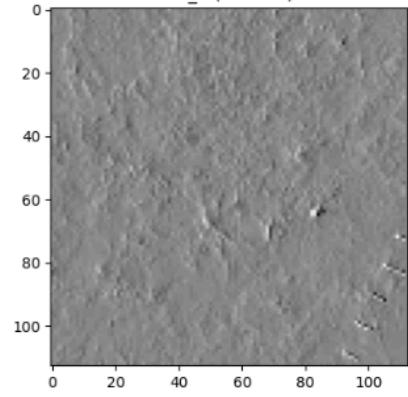
LL_2 (Approximation)



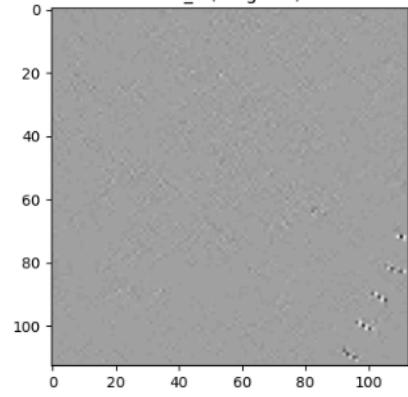
LH_2 (Horizontal)



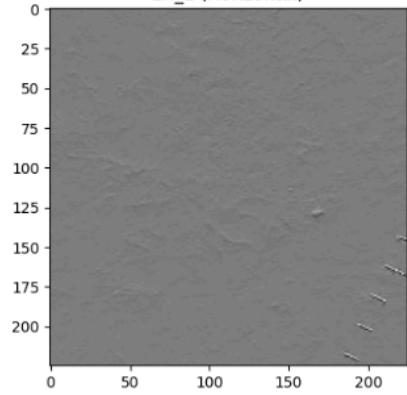
HL_2 (Vertical)



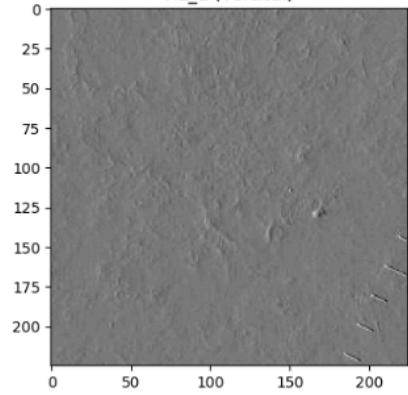
HH_2 (Diagonal)



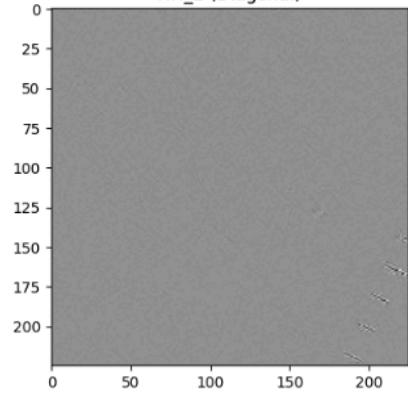
LH_1 (Horizontal)



HL_1 (Vertical)

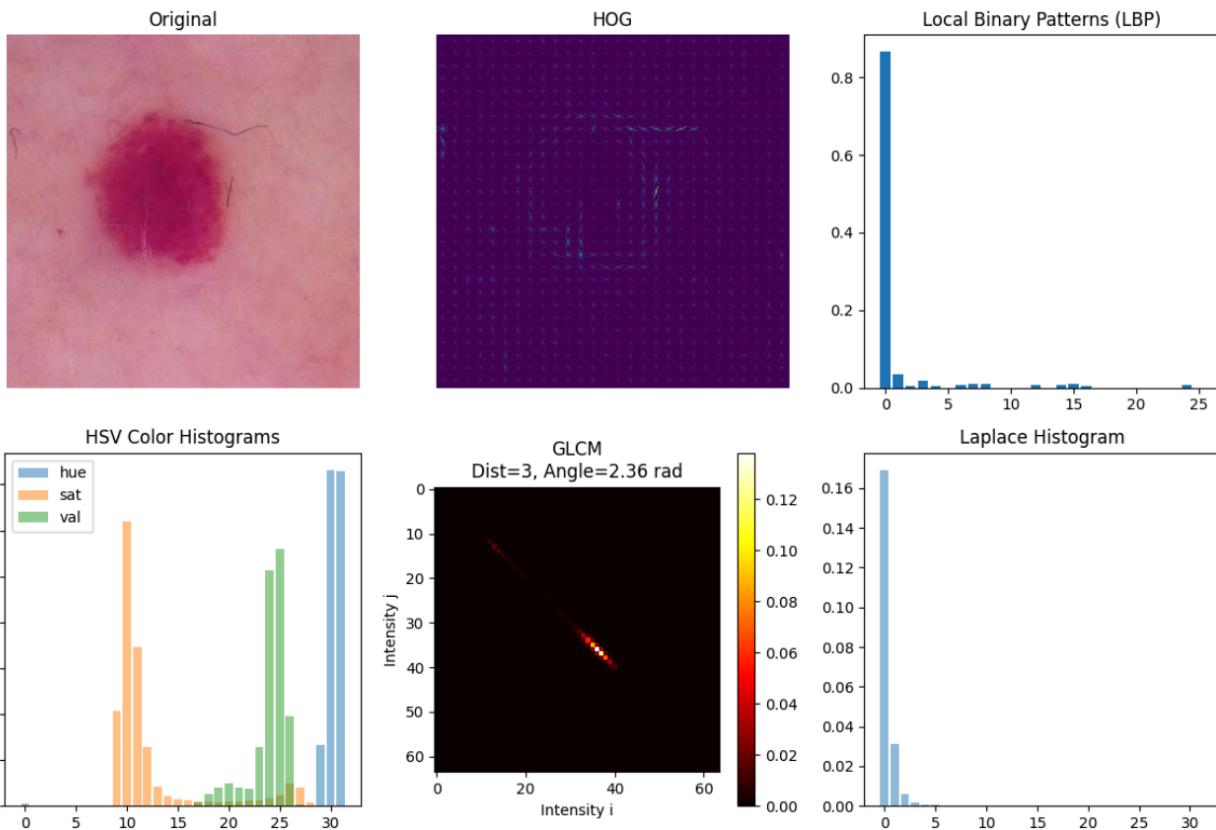


HH_1 (Diagonal)



VASC (Vascular lesion)

Image id: 6154
Class:VASC



HSV Metrics

	Mean	Std	Entropy	Contrast
h	168.756804	1.860798	0.489277	4.661649
s	196.234163	24.350305	0.785036	77.135967
v	164.716464	12.689486	0.713761	25.926955

GLCM Metrics

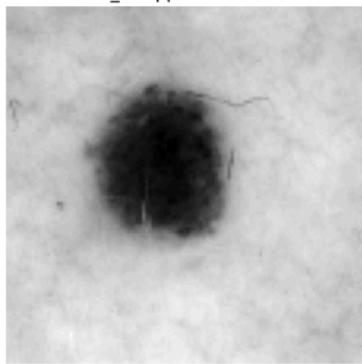
	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.660124	0.386805	0.823111	0.241315	0.993136	0.058233

Shape Metrics

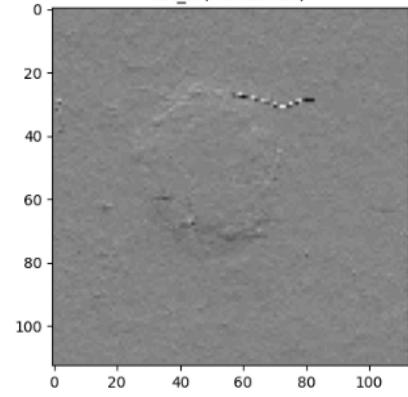
	circularity	eccentricity	convexity
6154	0.623895	0.475786	0.83546

Wavelet

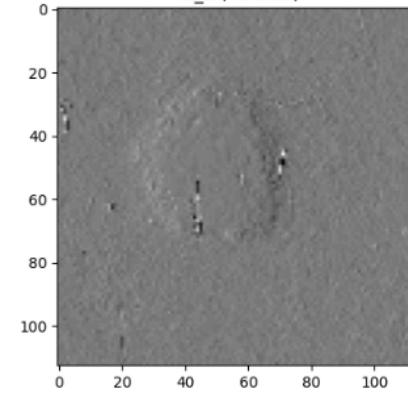
LL_2 (Approximation)



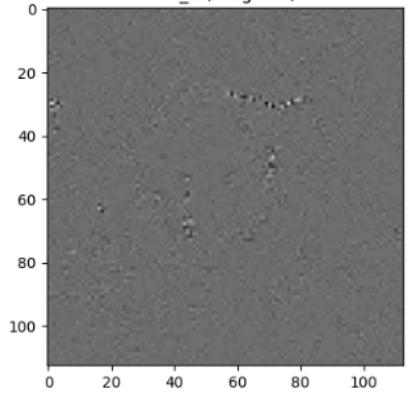
LH_2 (Horizontal)



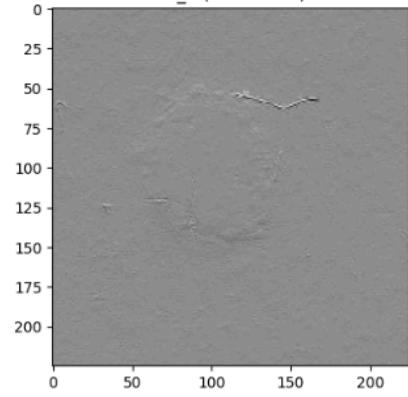
HL_2 (Vertical)



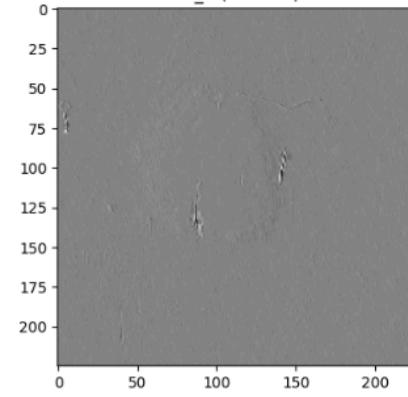
HH_2 (Diagonal)



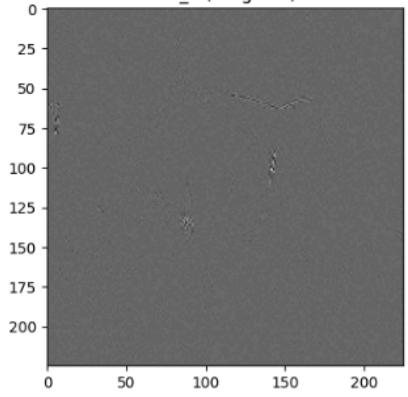
LH_1 (Horizontal)



HL_1 (Vertical)

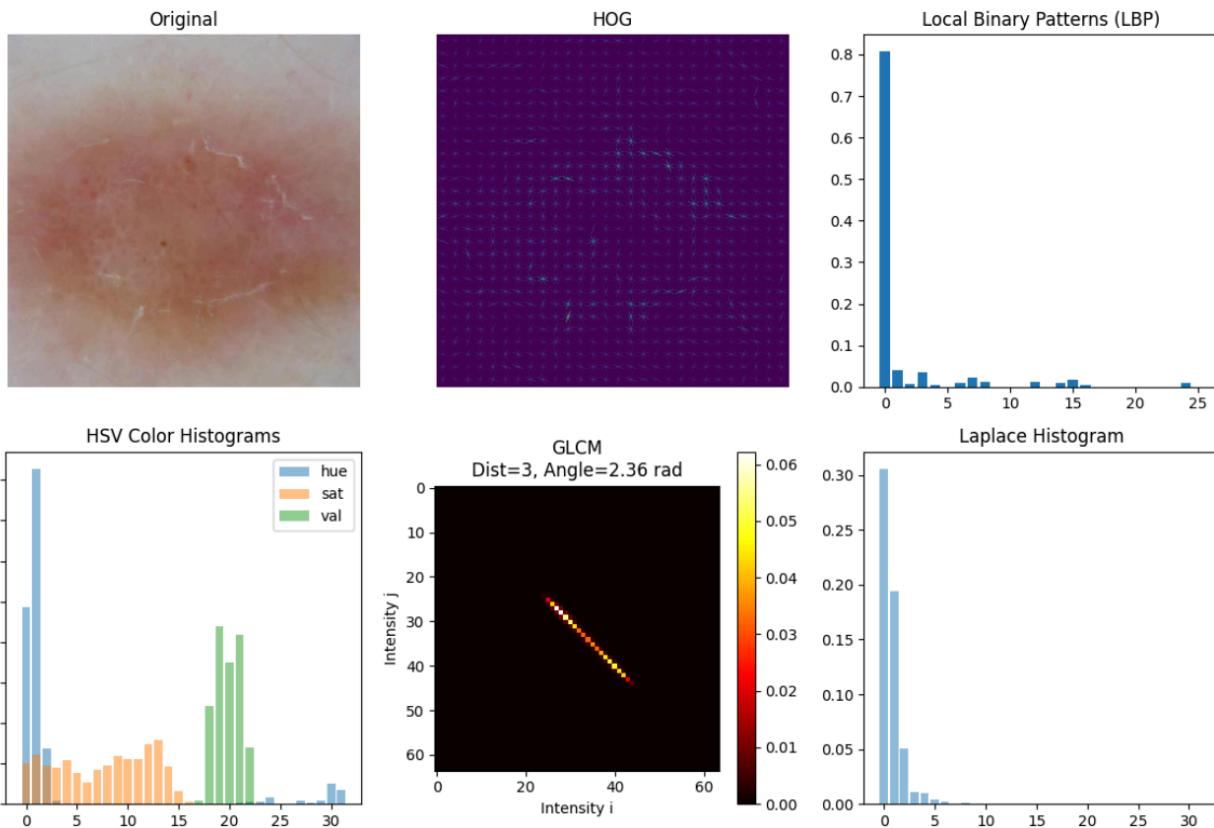


HH_1 (Diagonal)



DF (Dermatofibroma)

Image id: 7391
Class:DF



HSV Metrics

	Mean	Std	Entropy	Contrast
h	47.169488	68.467169	0.667929	40.416507
s	24.862004	18.141164	0.742340	44.960313
v	173.000000	4.476193	0.549389	10.481083

GLCM Metrics

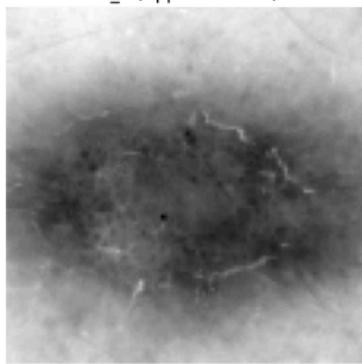
	contrast	dissimilarity	homogeneity	energy	correlation	ASM
0	0.228112	0.206503	0.898835	0.191714	0.996191	0.036754

Shape Metrics

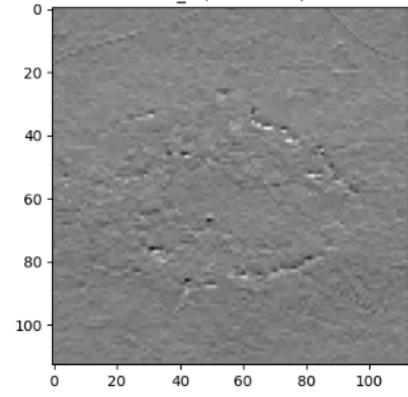
	circularity	eccentricity	convexity
7391	0.330621	0.964472	0.851269

Wavelet

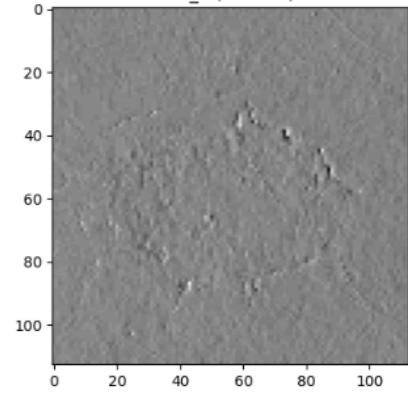
LL_2 (Approximation)



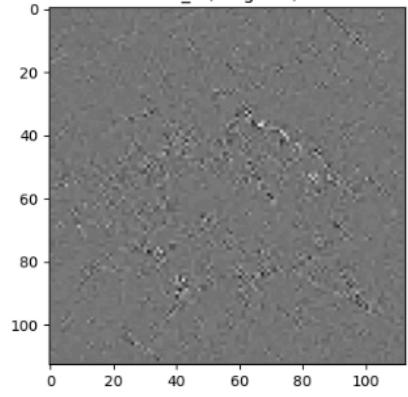
LH_2 (Horizontal)



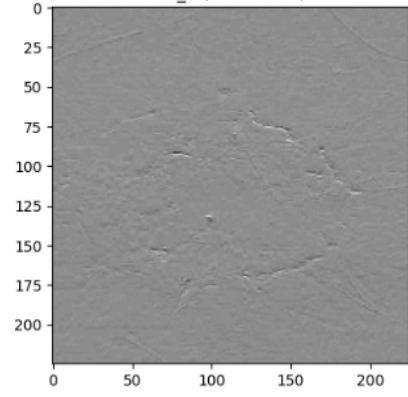
HL_2 (Vertical)



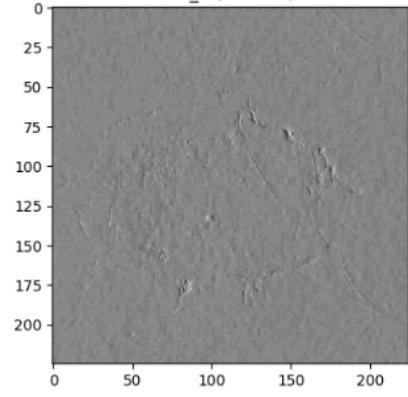
HH_2 (Diagonal)



LH_1 (Horizontal)



HL_1 (Vertical)



HH_1 (Diagonal)

