

UNIVERSITY OF BUCHAREST

FACULTY OF
MATHEMATICS AND
COMPUTER SCIENCE



ARTIFICIAL INTELLIGENCE SPECIALIZATION

PRACTICAL MACHINE LEARNING PROJECT

NAIL DISEASE DETECTION USING UNSUPERVISED MODELS

Student
Vîrtopeanu Sebastian-Filip

Scientific Coordinator
Prof. Dr. Radu Ionescu

Bucharest, January 2025

Contents

1	Introduction	4
1.1	Exploratory Data Analysis	4
1.1.1	Class distribution	4
1.1.2	Visual inspection of nail conditions	5
2	Exploratory Analysis on the Dataset	7
2.1	Datasets structure	7
2.2	HSV analysis	7
2.3	HOG analysis	8
2.4	Local Binary Pattern	9
3	Feature Representation	10
3.1	Preprocessing	10
3.2	Feature Set 1: HSV Bins	10
3.3	Feature Set 2: HOG	10
3.4	Feature Set 3: LBP	11
4	Models	12
4.1	DBSCAN	12
4.2	Agglomerative Hierarchical Clustering (AHC).	13
4.3	SVM baseline.	15
5	Results	16
5.1	Random baseline	16
5.2	DBScan	17
5.2.1	LBP	17
5.2.2	HSV	19
5.3	AHC	20
5.3.1	LBP	20
5.3.2	HSV	22
5.3.3	HOG	23

5.4	SVM	25
5.4.1	HSV	25
5.4.2	LBP	25
6	Conclusion	26

Chapter 1

Introduction

This project builds on a dataset published by Nikhil Gurav in a [Kaggle competition](#) and is used for the second project in the first year of the Artificial Intelligence Master's program. The objective is to identify whether the nails shown in an image exhibit a specific disease, from these predictions, we also derive a binary classification of **healthy** vs **unhealthy**. We present the dataset and conduct exploratory data analysis (EDA), then describe the methods we apply, including the hyperparameters the unsupervised models used and the results obtained.

1.1 Exploratory Data Analysis

1.1.1 Class distribution

Each split is further divided into six diagnostic categories. The number of images per class is reported below.

Training set.

- `Acral_Lentiginous_Melanoma`: 735 images
- `Healthy_Nail`: 323 images
- `Onychogryphosis`: 677 images
- `blue_finger`: 603 images
- `clubbing`: 767 images
- `pitting`: 639 images

Validation set.

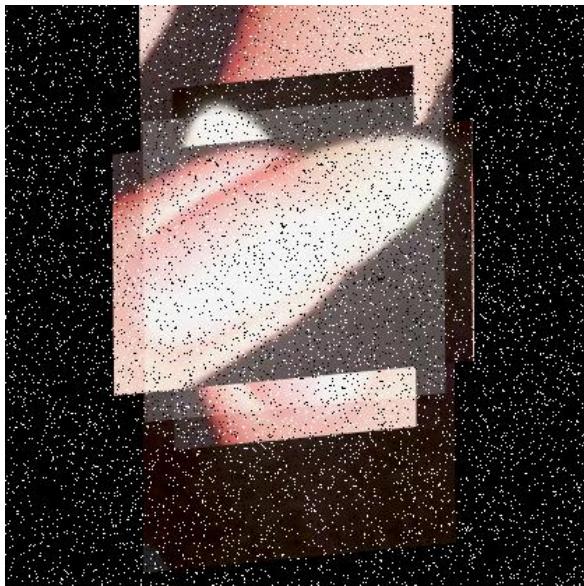
- `Acral_Lentiginous_Melanoma`: 18 images
- `Healthy_Nail`: 20 images
- `Onychogryphosis`: 12 images
- `blue_finger`: 9 images
- `clubbing`: 16 images
- `pitting`: 16 images

Overall, the training split is slightly imbalanced, `Healthy_Nail` contains fewer samples than the disease classes, while `clubbing` and `Acral_Lentiginous_Melanoma` are among the most represented. In the validation split, the number of samples per class is small, which may lead to higher variance in validation metrics.

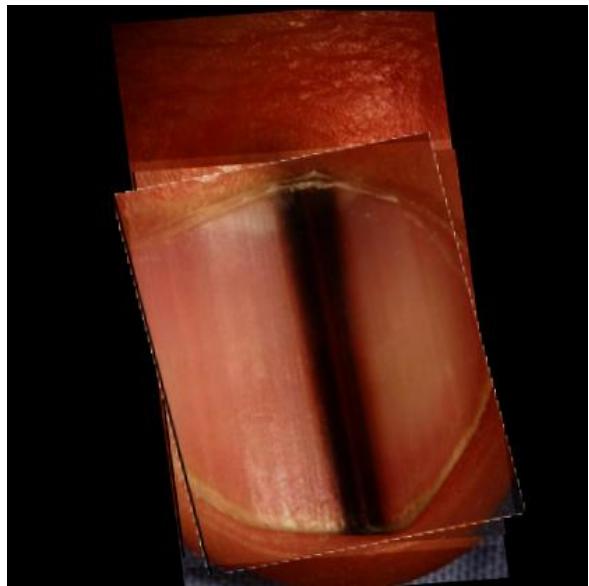
1.1.2 Visual inspection of nail conditions

To complement the overview, we perform a short visual inspection by sampling images from each class (Figures 1.1a - 1.1b). The categories can be distinguished by characteristic visual cues:

- **Healthy nail** (Figure 1.1a): smooth nail plate, regular color and a uniform surface texture.
- **Acral Lentiginous Melanoma** (Figure 1.1b): irregular dark pigmentation bands or patches, often asymmetric and with non-uniform intensity.
- **Onychogryphosis** (Figure 1.1c): markedly thickened nail plate with pronounced curvature and a rough, layered appearance.
- **Blue finger** (Figure 1.1d): bluish discoloration of the nail and surrounding tissue, suggestive of reduced oxygenation or vascular issues.
- **Clubbing** (Figure 1.1e): increased convexity and bulbous enlargement of the distal phalanx, with the nail appearing more rounded.
- **Pitting** (Figure 1.1f): small depressions on the nail surface, producing a dotted, uneven texture.



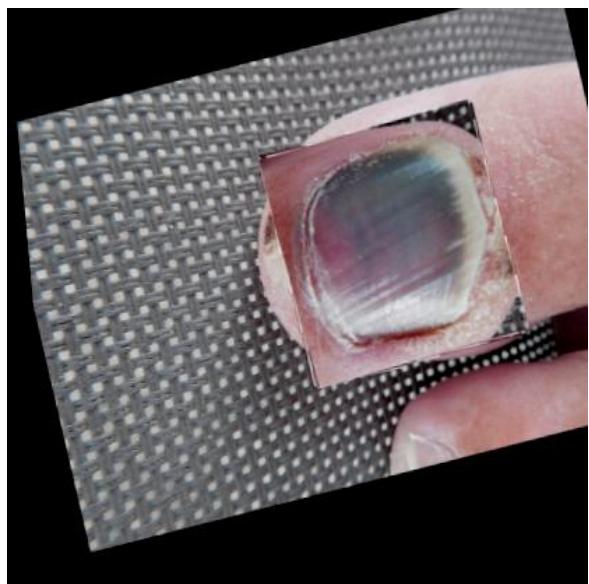
(a) Healthy nail.



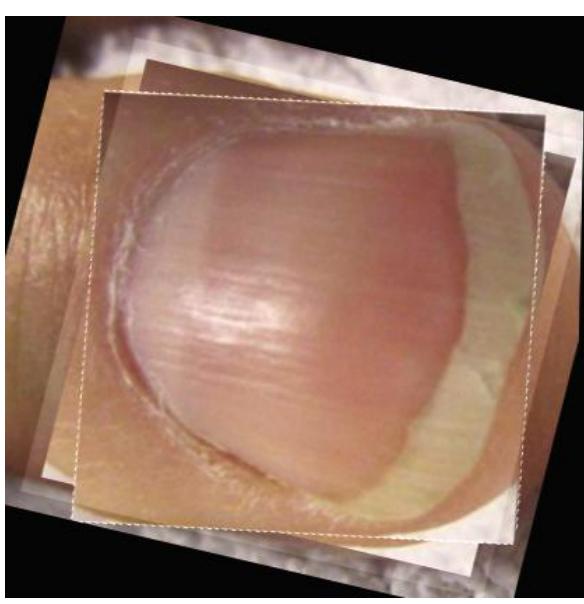
(b) Acral Lentiginous Melanoma.



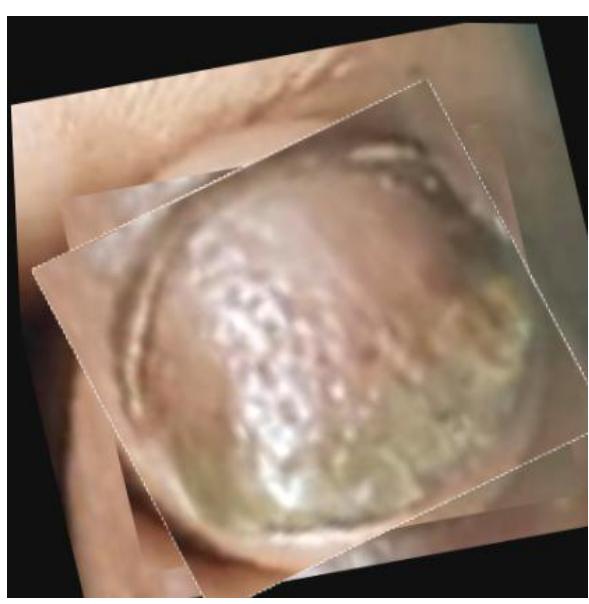
(c) Onychogryphosis.



(d) Blue finger.



(e) Clubbing.



(f) Pitting.

Figure 1.1: Representative samples from each nail condition class in the dataset.

Chapter 2

Exploratory Analysis on the Dataset

2.1 Datasets structure

The dataset is organized into folders corresponding to six classes: five nail diseases characterized by structural and color related changes and one *healthy* class. For the healthy samples, salt and pepper noise seems to have been added. In addition, the nail region appears to be highlighted by increased brightness, as shown in Fig. 1.1d.

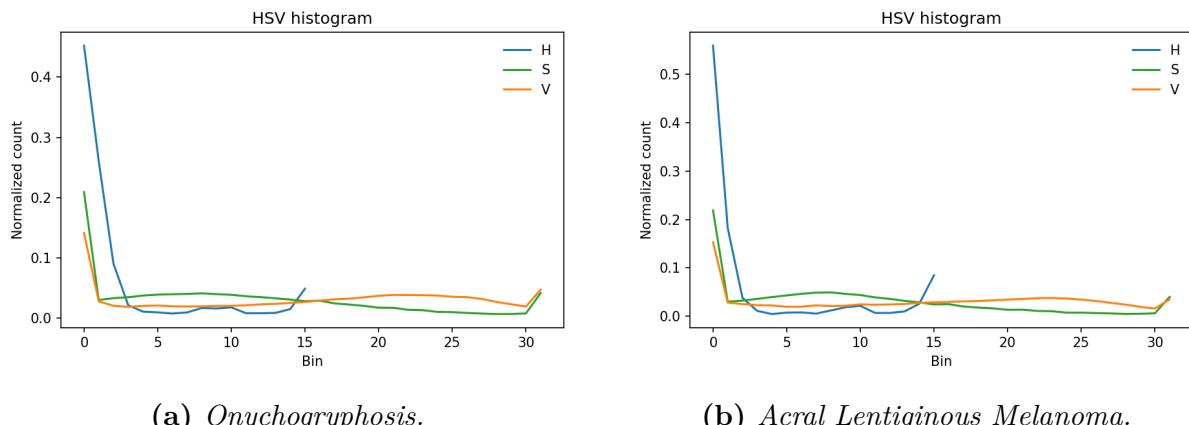


Figure 2.1: HSV channel histograms for two training classes.

2.2 HSV analysis

Since several nail diseases manifest primarily through color changes, we first convert the images from RGB to the HSV color space, which separates color information (Hue) from intensity components (Saturation and Value). As illustrated in Fig. 2.1a and Fig. 2.1b, the Saturation and Value distributions are largely similar across classes (a trend also observed for the remaining categories). Therefore, we focus on the Hue channel, which retains the most discriminative color cues and use its histogram bins to compute distances between

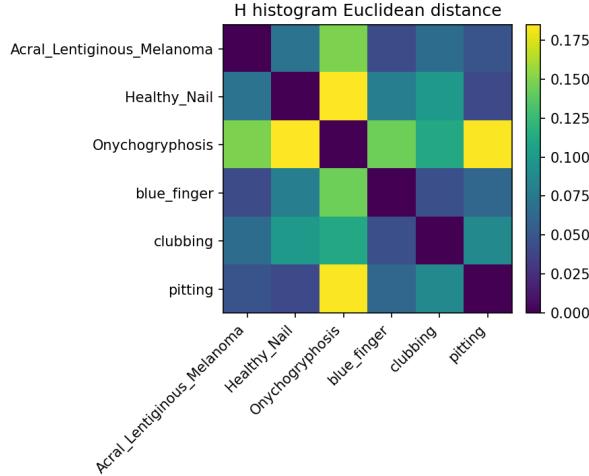


Figure 2.2: Pairwise distance matrix computed from HSV color histograms.

classes (Fig. 2.2).

2.3 HOG analysis

Beyond color driven details (already addressed through the HSV representation), we also examine shape information using a gradient based descriptor. Using these HOG feature vectors, we compute pairwise distances between classes and summarize them in a distance matrix. As shown in Fig. 2.3, the *healthy* class appears clearly separated from the pathological categories, indicating that gradient/textural patterns provide additional discriminative signal.

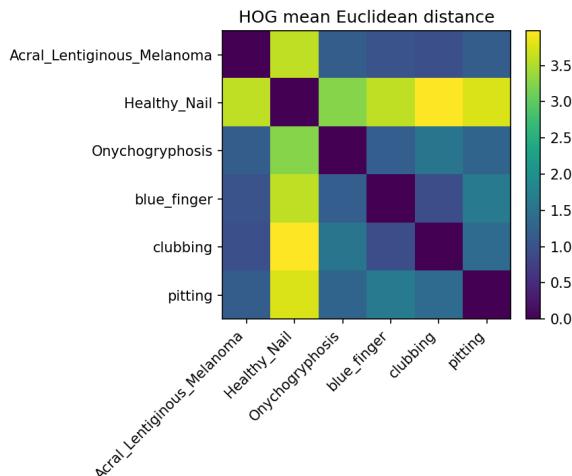


Figure 2.3: Pairwise distance matrix computed from HOG feature histograms.

2.4 Local Binary Pattern

Since certain conditions (*Onychogryphosis*) present distinctive surface textures, we also explore a pattern based descriptor: Local Binary Patterns. LBP operates on grayscale images and encodes local microtextures by thresholding each pixel against its neighborhood, creating a representation of repetitive patterns and surface irregularities.

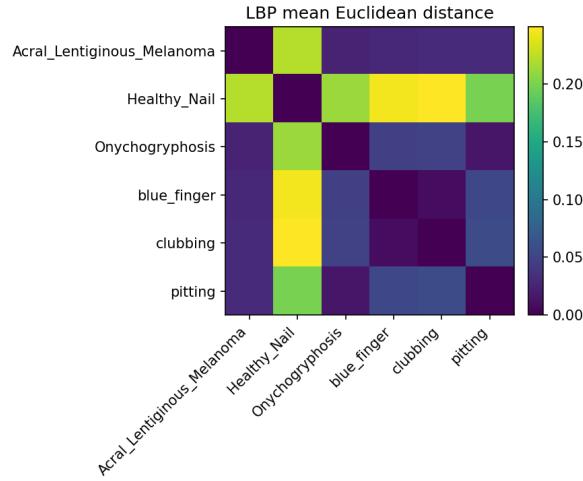


Figure 2.4: Pairwise distance matrix computed from LBP feature histograms.

Chapter 3

Feature Representation

To obtain informative groupings, we extract handcrafted descriptors for the cropped nail region. As discussed in Chapter 2, we consider three complementary feature families designed to capture color information, orientation structure and local texture patterns. The approach here is that we propose different sets of hyperparamater and in a grid search we will try to test them against our data.

3.1 Preprocessing

Because several of our descriptors are high dimensional, we can use **PCA** to reduce redundancy and improve computational efficiency (we will also use it to display the information in a 2D graph). In addition, since the clustering methods we consider are distance based, we need to standardize all features so that each dimension contributes on a comparable scale to the distance computation. These preprocessing steps are applied consistently across all feature, model combinations, also the PCA dimensionality was treated as a tunable hyperparameter and included in our grid search.

3.2 Feature Set 1: HSV Bins

For this feature set, we represent each image using HSV color histograms, where the number of bins per channel is treated as a feature specific hyperparameter. We evaluate bin counts in $\{16, 32, 64\}$.

3.3 Feature Set 2: HOG

For the HOG representation, each cropped nail image is resized to a fixed resolution and described through local gradient orientation.

For an image resized to 128×128 with `pixels_per_cell`(8, 8), `cells_per_block`(2, 2) and `orientations`=9, we obtain a 16×16 grid of cells. Using overlapping 2×2 blocks returns 15×15 blocks and each block contributes $2 \times 2 \times 9 = 36$ values. Therefore, the final descriptor size is $15 \times 15 \times 36 = 8100$ features per image. 8100 dimensions seems to be too much for our image, comparing Fig. 3.1a with Fig. 3.1c, we can see a clear differences of patterns. Using a configuration with `pixels_per_cell`(16, 16) at 128×128 , which returns an 8×8 cell grid, 7×7 blocks resulting in $7 \times 7 \times 36 = 1764$ features per image (Fig. 3.1).

DBSCAN and AHC are evaluated using the same clustering grids as in Feature Set 1. For the supervised baseline, we train a `LinearSVC` with default $C = 1$ and `max_iter`=5000. All HOG vectors are standardized prior to modeling.

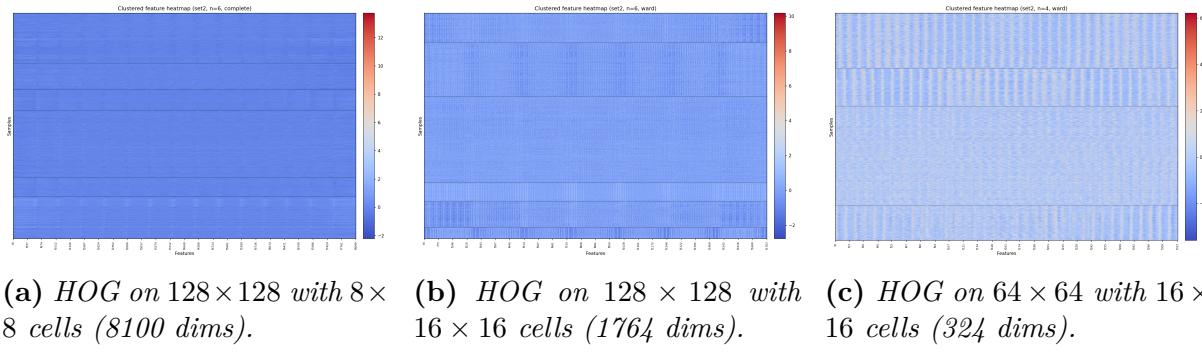


Figure 3.1: *HOG descriptor visualizations for different resize resolutions and cell sizes after AHC clustering.*

3.4 Feature Set 3: LBP

To capture local texture patterns, we extract Local Binary Pattern (LBP) descriptors from grayscale nail images resized to 128×128 . Unlike the previous fixed configuration, we treat the LBP neighborhood size as a feature specific hyperparameter and grid search `radius` $\in\{1,2,3\}$, `n_points` $\in\{8,16,24\}$.

Chapter 4

Models

We evaluate two unsupervised, distance based clustering methods DBSCAN and Agglomerative Hierarchical Clustering (AHC) and compare them against a supervised baseline using a Support Vector Machine (SVM). Since the dataset contains six original classes and the images are relatively noisy or difficult to detect, we report results in two settings: a multiclass formulation (six classes) and a binary formulation that groups samples into *healthy* vs. *unhealthy*.

Because DBSCAN relies on a distance measure, we use the elbow method to select appropriate distance related hyperparameters (ε for DBSCAN).

4.1 DBSCAN

DBSCAN struggled to separate the dataset into all six disease specific categories, in the multiclass setting, the best clustering to label agreement reached only around $\sim 20\%$. In contrast, DBSCAN performed substantially better in the binary setting (*healthy* vs. *unhealthy*), achieving up to 75% accuracy on the train dataset (including noise that was label wrong) for certain hyperparameter configurations. This behavior is consistent with the structure observed in Fig. 4.1, where the algorithm predominantly forms two clusters, effectively collapsing the problem into a healthy/non-healthy partition.

Its behavior is highly sensitive to the distance hyperparameters: if ε is too small, most samples are labeled as noise (Fig. 4.2a), whereas if ε is too large, distinct groups may collapse into a single dominant cluster (Fig. 4.2b). To guide the choice of ε , we rely on k -distance elbow plots and evaluate a small grid of candidate values for each feature representation.

From the elbow location, we observed that the HSV histogram features suggest $\varepsilon \approx 4.20$ (Fig. 4.3a), HOG features suggest $\varepsilon \approx 4.10$ (Fig. 4.3b) and LBP features suggest $\varepsilon \approx 1.00$ (Fig. 4.3c), but that is only a single test case because the number might change if we increase the number of neighbors. Accordingly, we test ε values in a narrow neighborhood

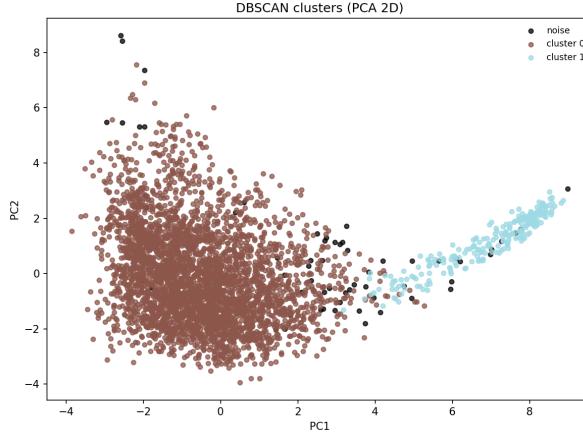
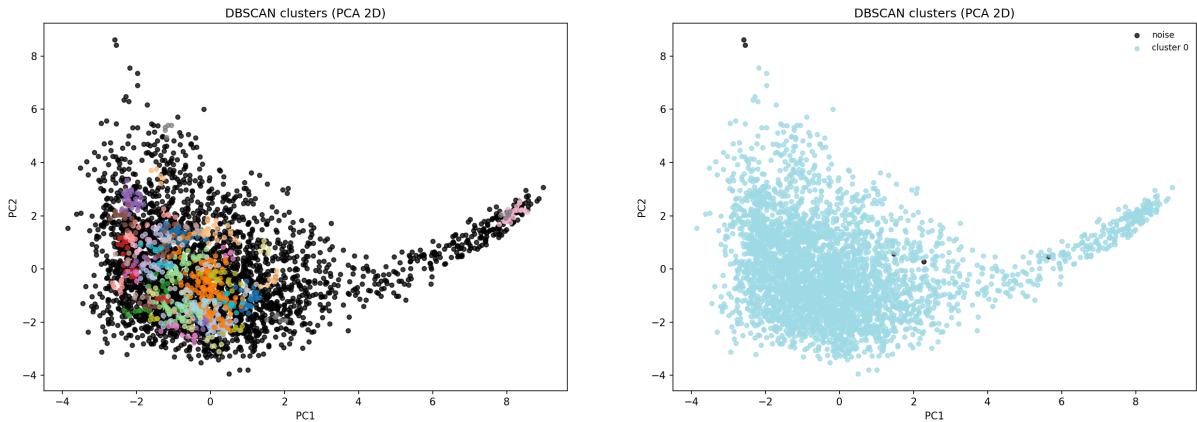


Figure 4.1: DBSCAN result on the training set, the blue cluster corresponding to healthy vs. unhealthy, corresponding to brown.

around these knees ($\{4.00, 4.20, 4.40, 4.60\}$ for HSV, $\{3.90, 4.10, 4.30, 4.50\}$ for HOG and $\{0.80, 1.00, 1.20, 1.40\}$ for LBP).

Regarding `min_samples`, we set `min_samples= 3` as a default *permissive* configuration to reduce the risk of labeling too many points as noise and create an initial intuition about the distances, we evaluate stricter settings $\text{min_samples} \in \{5, 8, 10\}$, for each `min_samples` value we recompute the corresponding k -distance plot (with $k = \text{min_samples}$) and select ε near the knee.



(a) Small ε , most samples are labeled as noise (-1).

(b) Large ε , most samples collapse into one cluster.

Figure 4.2: Failure modes of DBSCAN.

4.2 Agglomerative Hierarchical Clustering (AHC).

AHC constructs a hierarchical clustering by iteratively merging the most similar clusters according to a chosen linkage rule (single, complete or average linkage). Two hyperparameters that we explored are the *distance metric* and the *linkage criterion*. In our

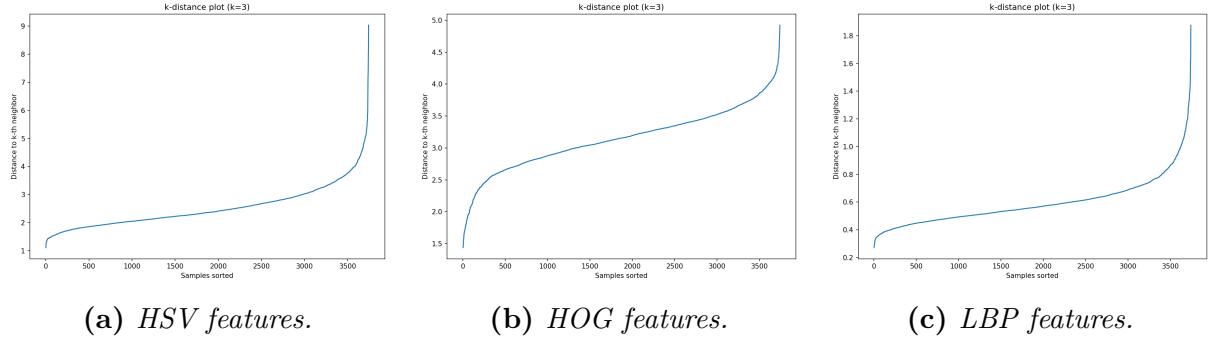


Figure 4.3: *k*-distance plots used to estimate a suitable ε for DBSCAN across different feature representations.

experiments, we primarily use Euclidean distance, because we have continuous feature vectors and Manhattan distance, which can provide sharper differences. Although cosine distance is often useful for semantic embeddings, we did not use it because we encode low-level visual statistics rather than semantic similarity.

To interpret and justify the AHC behavior, we visualize the feature matrices as heatmaps, where rows correspond to samples and columns correspond to features. Starting from the raw feature matrix (Fig. 4.4a), we already observe coherent blocks of similar patterns, with the healthy class forming one of the most apparent regions. After reordering samples according to the hierarchical clustering (Fig. 4.4b), these blocks become more clearly separated. We include representative clustering examples for each feature representation (LBP, HSV and HOG) in Fig. 4.5c - Fig. 4.5b, with the LBP descriptor being the most separable graph.

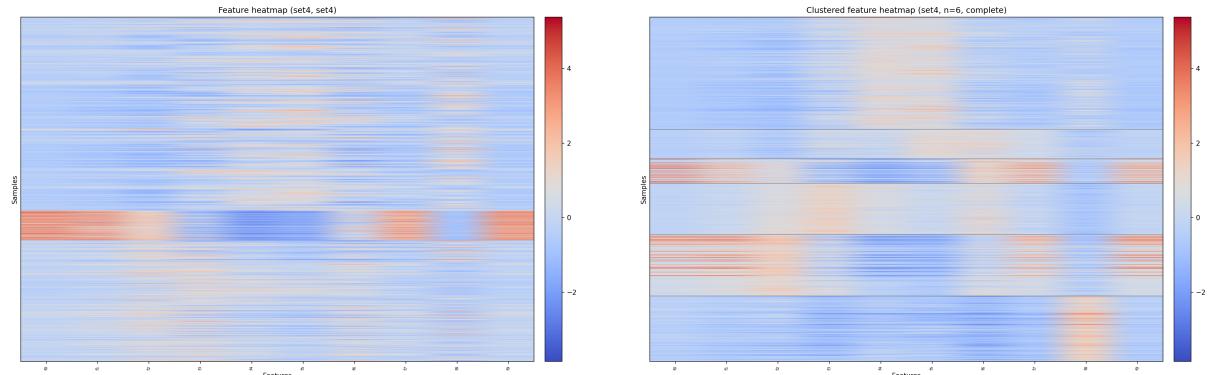


Figure 4.4: Heatmaps of LBP features.

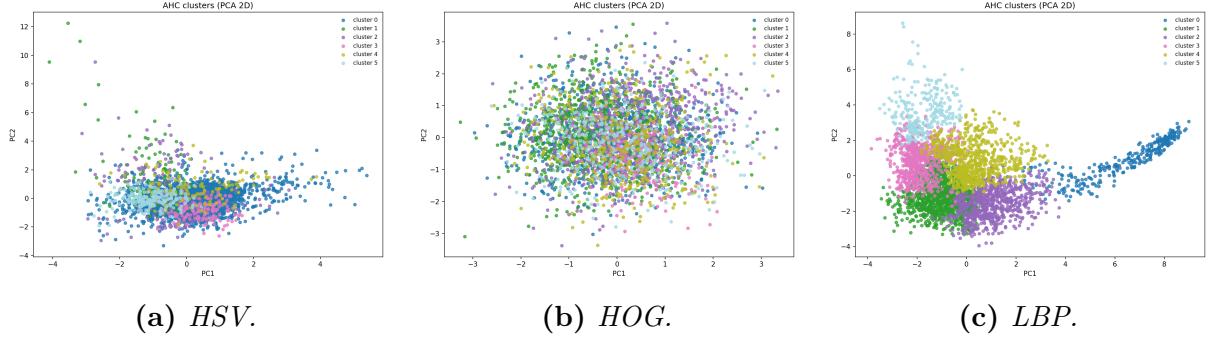


Figure 4.5: AHC clustering examples across feature representations.

4.3 SVM baseline.

As a supervised reference, we train an SVM classifier. We chose SVMs as a strong candidate for this task because they can separate binary classes by default and for multi-class it is adopting a one-vs-all strategy, similar with what we want to explore in the unsupervised methods.

Chapter 5

Results

To evaluate the results, we first remap the cluster IDs to the true labels in a way that maximizes accuracy, since cluster labels are arbitrary. For example, the model may perfectly identify the *Healthy nail* class, but if it assigns it a different cluster ID, the raw accuracy would appear low even though the grouping is correct. After this optimal label alignment, we report the accuracy and we also compute the ARI and NMI scores.

In order to create a common abbreviation dictionary, we will use the following abbreviations for our classes:

- **H** = Healthy_Nail
- **ALM** = Acral_Lentiginous_Melanoma
- **ONY** = Onychogryphosis
- **BF** = blue_finger
- **CLB** = clubbing
- **PIT** = pitting

5.1 Random baseline

With the random baseline (50 runs on the validation split, 91 samples, 6 classes), performance is low in the multiclass setting: the mean overall accuracy is $\sim 0.16 \pm 0.04$, which is close to what you would expect from guessing among 6 labels.

In the binary setting (healthy vs. unhealthy), the mean accuracy is much higher at $\sim 0.62 \pm 0.04$, but this is largely explained by class imbalance: there are 71 unhealthy samples versus 20 healthy.

Tabel 5.1: Random baseline performance on the validation split.

Metric	Mean
Overall accuracy (6-class)	0.16
Binary accuracy (healthy vs. unhealthy)	0.62

Tabel 5.2: Per class accuracy for the random baseline on the validation split.

Class	Count	Accuracy mean
ALM	18	0.15
H	20	0.16
ONY	12	0.17
BF	9	0.17
CLB	16	0.17
PIT	16	0.16

5.2 DBScan

5.2.1 LBP

In our DBSCAN experiments, separating all six disease categories was consistently difficult, as samples tend to form sparse and partially overlapping regions in feature space. Our most successful results were obtained with the LBP feature representation, where DBSCAN achieved strong performance for the binary task (*healthy* vs. *unhealthy*). As shown in Fig. 5.1a, the clustering largely separates non-healthy nails (brown) from a compact group containing most healthy samples (light blue). This outcome is also reflected in Table 5.3, where one cluster is perfectly pure for *Healthy_Nail*. Still, some healthy nails are labeled as noise.

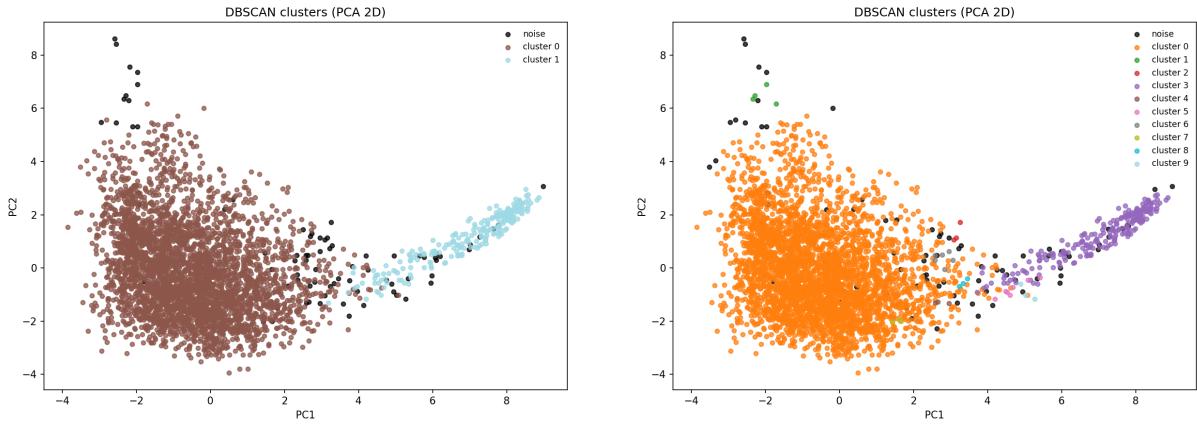
However, when we attempt a finer grained multiclass mapping, the clustering becomes less interpretable: entropy increases, dominant clusters mix several disease labels and additional small clusters appear (Fig. 5.1b). We explored stricter density constraints by increasing `min_samples`, but this primarily expanded the noise cluster rather than improving class level separability present in Table 5.4.

Tabel 5.3: DBSCAN cluster composition for an LBP configuration.

Cluster	Size	Entropy	Purity	Top-1	Top-2	Top-3
-1	81	1.48	0.65	H (53)	ALM (15)	ONY (7)
0	3432	2.38	0.22	CLB (767)	ALM (720)	ONY (670)
1	231	0.00	1.00	H (231)	–	–

Tabel 5.4: DBSCAN cluster composition using LBP features in a multiclass attempt.

Cluster	Size	Entropy	Purity	Top-1	Top-2	Top-3
-1	379	2.16	0.47	H (180)	ALM (68)	BF (48)
0	3229	2.35	0.23	CLB (743)	ALM (667)	ONY (643)
1	8	0.00	1.00	BF (8)	—	—
2	125	0.00	1.00	H (125)	—	—
3	3	0.92	0.67	ONY (2)	PIT (1)	—



(a) DBSCAN attempt for binary classification with LBP.

(b) DBSCAN attempt for multiclass classification with LBP.

Figure 5.1: DBSCAN visualizations using LBP features: (a) configuration summarized in Table 5.3 and (b) configuration summarized in Table 5.4.

Tabel 5.5: DBSCAN results with LBP features on the training set.

eps	min samples	clusters	noise	noise ratio	accuracy	clustered accuracy	accuracy	binary
0.50	12	1	3731	1.00	0.31	0.00	0.00	
0.50	8	13	3586	0.96	0.26	0.01	0.05	
0.50	5	68	2923	0.78	0.14	0.03	0.21	
0.50	3	162	2119	0.57	0.16	0.07	0.43	
0.67	12	4	1752	0.47	0.23	0.12	0.78	
0.67	3	38	565	0.15	0.25	0.21	0.85	
0.83	3	15	189	0.05	0.26	0.25	0.94	
1.17	3	7	31	0.01	0.27	0.27	0.98	
1.17	5	2	62	0.02	0.27	0.27	0.84	
1.00	3	10	84	0.02	0.27	0.27	0.97	
1.00	5	3	127	0.03	0.27	0.26	0.92	
1.33	3	5	7	0.00	0.21	0.21	0.91	

Tabel 5.6: DBSCAN results with LBP features on the validation dataset.

lbp	radius	eps	min samples	clusters	noise	noise ratio	accuracy	clustered accuracy	accuracy	binary
1	0.50		12	1	3731	1.00	0.00	0.00	0.00	
1	1.00		3	10	84	0.02	0.97	0.26	0.63	
1	1.17		3	7	31	0.01	0.99	0.32	0.74	
0.50	1.17		5	2	62	0.02	0.98	0.15	0.52	
2	1.00		5	3	127	0.03	0.97	0.16	0.61	

5.2.2 HSV

In contrast to the LBP representation, using HSV histogram bins with DBSCAN was considerably less successful. We attribute this primarily to the black padding present around many cropped nail images, which introduces a dominant dark region that can distort color histograms and reduce their discriminative power.

The first failure mode is illustrated in Fig. 5.2a and quantified in Table 5.9, where DBSCAN generates a large number of clusters (including many tiny ones). Although some of these small clusters are pure, the overall structure is unstable and does not align well with the six-class ground truth. The second failure mode is shown in Fig. 5.2b and Table 5.10, where the solution effectively degenerates into two groups (one large mixed cluster and noise), with high entropy and low purity in the dominant cluster.

Tabel 5.7: DBSCAN results with HSV features on the training dataset.

bins	eps	min samples	clusters	noise	noise ratio	accuracy	clustered	accuracy	accuracy	binary
16	0.50	3	0	3744	1.00	0.00	0.00	0.00	0.00	
16	2.00	3	78	2257	0.60	0.18	0.40	0.37		
64	2.00	3	53	2557	0.68	0.19	0.32	0.30		
16	2.00	5	18	2659	0.71	0.21	0.29	0.23		
32	2.00	3	58	2679	0.72	0.19	0.28	0.26		
16	1.83	3	66	2749	0.73	0.18	0.27	0.26		
64	2.00	5	16	2927	0.78	0.23	0.22	0.24		
16	1.67	5	17	3452	0.92	0.28	0.08	0.02		
16	1.17	3	5	3726	1.00	0.50	0.00	0.01		
64	1.83	12	2	3716	0.99	0.43	0.01	0.01		
64	1.17	3	1	3741	1.00	1.00	0.00	0.00		
16	1.00	12	0	3744	1.00	0.00	0.00	0.00		

Tabel 5.8: DBSCAN results with HSV features on the validation dataset.

bins	eps	min samples	clusters	noise	noise ratio	accuracy	clustered	accuracy	accuracy	binary
16	0.50	3	0	3744	1.00	0.00	0.00	0.00	0.00	
16	2.00	3	78	2257	0.60	0.40	0.35	0.20		
64	2.00	3	53	2557	0.68	0.32	0.28	0.15		
16	2.00	5	18	2659	0.71	0.29	0.26	0.16		
32	2.00	3	58	2679	0.72	0.28	0.25	0.11		
16	1.83	3	66	2749	0.73	0.27	0.24	0.10		

Tabel 5.9: DBSCAN cluster composition for an HSV configuration and a smaller ϵ .

Cluster	Size	Entropy	Purity	Top-1	Top-2	Top-3
-1	2089	2.38	0.25	CLB (523)	ALM (482)	ONY (403)
0	1442	2.52	0.26	PIT (377)	H (271)	ONY (254)
1	3	0.00	1.00	ALM (3)	–	–
2	8	1.30	0.63	BF (5)	ALM (2)	CLB (1)
3	3	0.92	0.67	ALM (2)	CLB (1)	–
4	4	1.50	0.50	ALM (2)	BF (1)	CLB (1)
56	3	0.00	1.00	PIT (3)	–	–
57	3	0.92	0.67	PIT (2)	CLB (1)	–

Tabel 5.10: DBSCAN cluster composition for an HSV configuration and a larger ϵ .

Cluster	Size	Entropy	Purity	Top-1	Top-2	Top-3
-1	488	2.30	0.28	CLB (137)	ALM (121)	BF (111)
0	3256	2.55	0.19	CLB (630)	ALM (614)	PIT (605)

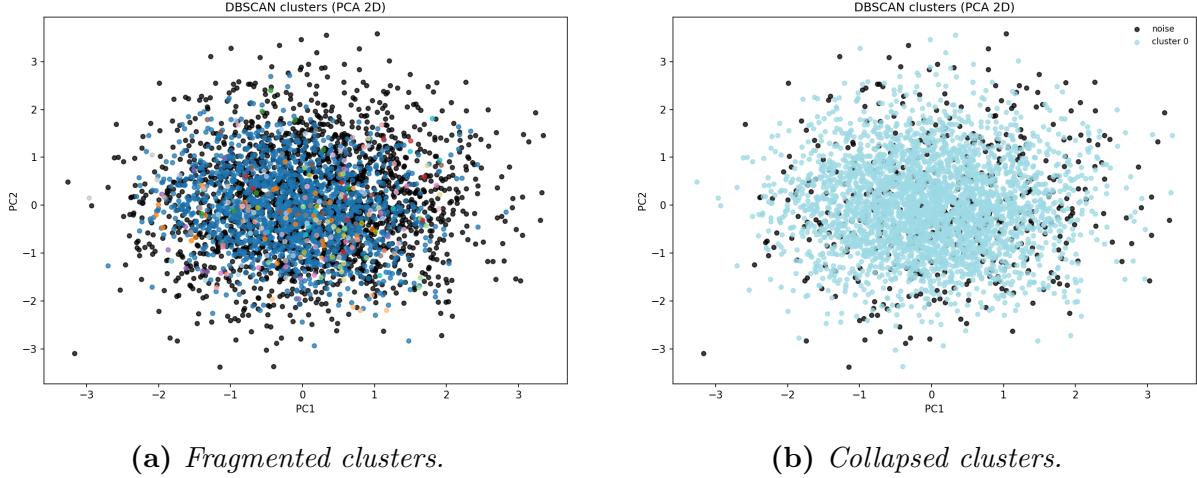


Figure 5.2: DBSCAN failure modes using HSV histogram features: (a) configuration summarized in Table 5.9 and (b) configuration summarized in Table 5.10.

5.3 AHC

5.3.1 LBP

Compared to DBSCAN, AHC produced noticeably cleaner multiclass structure in our visualizations. While DBSCAN frequently assigns a large fraction of samples to the noise label, AHC always returns a complete partition of the dataset. As shown in Fig. 5.3, several classes form well defined clusters, with the *Healthy_Nail* group in particular appearing highly compact and separable.

This qualitative behavior is supported by the cluster statistics in Table 5.11. In particular, cluster 0 is strongly dominated by *Healthy_Nail* (high purity, low entropy). The main limitation is that certain disease categories especially *Acral_Lentiginous_Melanoma*, *clubbing* and *pitting* remain partially mixed across multiple clusters, as reflected by higher entropies and lower purities for clusters 1-5. This overlap is plausible given that these conditions can manifest as subtle edge changes, discoloration or small surface malformations, which are harder to separate using low level handcrafted descriptors alone.

Tabel 5.11: AHC cluster composition with LBP features.

Cluster	Size	Entropy	Purity	Top-1	Top-2	Top-3
0	296	0.49	0.93	H (276)	ONY (7)	PIT (6)
1	1103	2.32	0.26	ALM (286)	CLB (269)	BF (219)
2	711	2.16	0.35	PIT (251)	ONY (242)	ALM (79)
3	690	1.93	0.40	CLB (273)	BF (221)	ALM (137)
4	695	2.27	0.27	ALM (190)	PIT (186)	ONY (168)
5	249	2.31	0.33	CLB (82)	BF (45)	ONY (44)

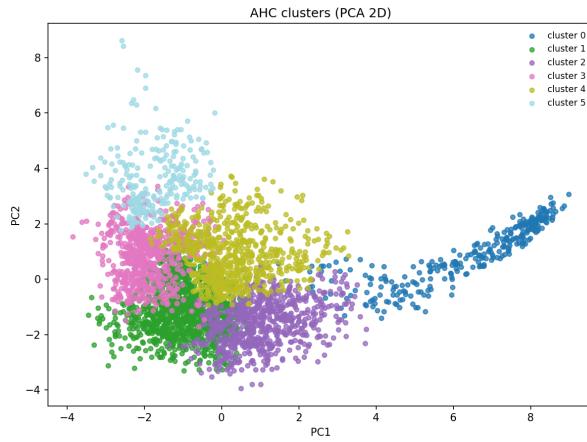


Figure 5.3: AHC clusters with LBP features.

Tabel 5.12: AHC results with LBP features on the training dataset.

n clusters	linkage	metric	accuracy	accuracy binary	ARI	NMI
4	WARD	EUCLIDEAN	0.36		0.98	0.12 0.20
4	AVERAGE	MANHATTAN	0.31		0.91	0.07 0.12
4	COMPLETE	MANHATTAN	0.31		0.91	0.09 0.13
6	WARD	EUCLIDEAN	0.35		0.98	0.11 0.19
6	AVERAGE	EUCLIDEAN	0.27		0.98	0.06 0.18
6	COMPLETE	MANHATTAN	0.30		0.91	0.08 0.13
8	WARD	EUCLIDEAN	0.31		0.98	0.10 0.18
8	AVERAGE	EUCLIDEAN	0.26		0.98	0.05 0.18
8	AVERAGE	MANHATTAN	0.33		0.91	0.09 0.13
8	COMPLETE	EUCLIDEAN	0.33		0.97	0.09 0.18
10	WARD	EUCLIDEAN	0.26		0.98	0.08 0.18
10	AVERAGE	MANHATTAN	0.33		0.91	0.09 0.13

Tabel 5.13: AHC results with LBP features on the validation dataset.

n clusters	linkage	metric	accuracy	accuracy binary	ARI	NMI
6	AVERAGE	MANHATTAN	0.62		0.91	0.40 0.49
8	AVERAGE	MANHATTAN	0.60		0.92	0.40 0.51
4	WARD	EUCLIDEAN	0.57		0.97	0.31 0.46
6	COMPLETE	MANHATTAN	0.57		0.92	0.33 0.41
10	AVERAGE	MANHATTAN	0.57		0.92	0.41 0.51

5.3.2 HSV

When using the HSV histogram representation, the AHC visualization still reveals some good block structure however, the separation between groups is noticeably weaker than for LBP. In particular, the PCA projection contains several apparent outliers, which makes the cluster boundaries less stable and complicates the interpretation (see Fig. 5.4a). Consistent with this observation, the label mapping indicates that HSV bins do not provide a clean partition even for the *Healthy_Nail* class: no cluster is dominated by healthy samples and the largest cluster mixes multiple disease labels with high entropy and low purity.

Tabel 5.14: *AHC using HSV features on the training dataset.*

bins	n clusters	linkage	metric	accuracy	ARI	NMI
32	4	AVERAGE	MANHATTAN	0.25	0.01	0.03
64	4	AVERAGE	MANHATTAN	0.24	0.01	0.02
64	6	AVERAGE	MANHATTAN	0.24	0.01	0.03
32	6	AVERAGE	MANHATTAN	0.23	0.01	0.03
32	4	COMPLETE	MANHATTAN	0.23	0.01	0.02
16	4	COMPLETE	MANHATTAN	0.23	0.01	0.01
64	10	WARD	EUCLIDEAN	0.22	0.01	0.03
32	10	WARD	EUCLIDEAN	0.20	0.01	0.04
32	8	WARD	EUCLIDEAN	0.21	0.00	0.03
64	10	AVERAGE	MANHATTAN	0.18	0.02	0.03
64	10	COMPLETE	MANHATTAN	0.17	0.01	0.02
16	8	AVERAGE	MANHATTAN	0.19	0.01	0.02

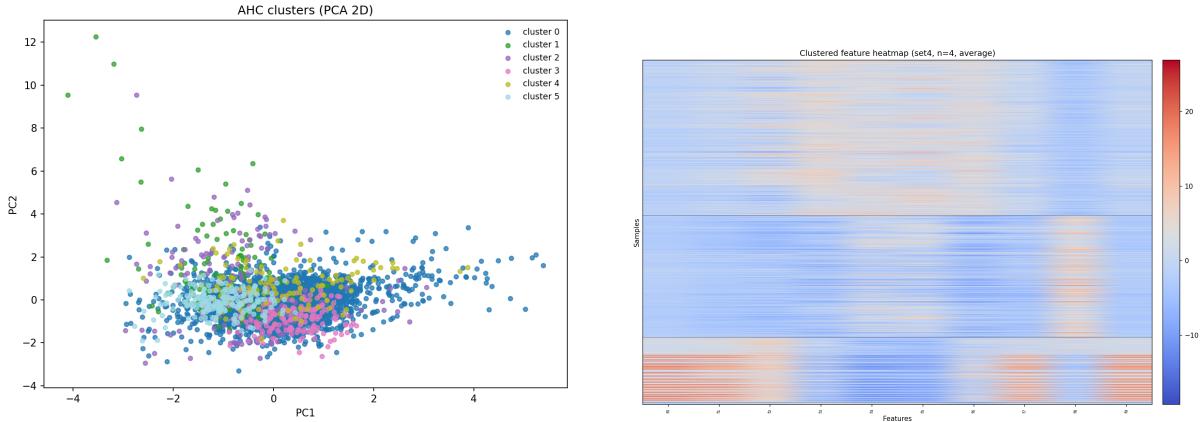
Tabel 5.15: *AHC using HSV features on the validation dataset.*

n clusters	linkage	metric	accuracy	accuracy binary	ARI	NMI
6	AVERAGE	MANHATTAN	0.41	0.78	0.11	0.23
6	WARD	EUCLIDEAN	0.38	0.79	0.09	0.21
6	AVERAGE	MANHATTAN	0.38	0.78	0.10	0.25
6	COMPLETE	MANHATTAN	0.38	0.81	0.10	0.22
8	AVERAGE	MANHATTAN	0.37	0.82	0.11	0.27

These limitations are summarized in Table 5.16. Cluster 0 contains the majority of samples but is highly contaminated (purity ≈ 0.21), with *Acral_Lentiginous_Melanoma*, *clubbing* and *pitting* all appearing as frequent labels. The remaining clusters are substantially smaller and only moderately purer (purity typically below ~ 0.47), which suggests that HSV color statistics alone are insufficient to reliably separate the six categories in our dataset.

Table 5.16: AHC cluster composition using HSV histogram features (multiclass setting).

Cluster	Size	Ent.	Pur.	Top-1	Top-2	Top-3
0	2955	2.55	0.21	ALM (625)	CLB (556)	PIT (546)
1	129	2.02	0.45	ONY (58)	CLB (33)	ALM (18)
2	149	2.14	0.38	CLB (57)	ONY (30)	PIT (27)
3	180	2.50	0.24	ONY (43)	CLB (39)	BF (35)
4	122	1.99	0.47	BF (57)	CLB (26)	ALM (17)
5	209	2.47	0.27	CLB (56)	BF (54)	ONY (27)



(a) AHC visualization for HSV features summarized in Table 5.16.

(b) Heatmap of HSV feature vectors reordered by the AHC dendrogram.

Figure 5.4: AHC using HSV histogram features: (a) clustering visualization and (b) reordered feature heatmap.

5.3.3 HOG

The HOG representation performs weakest in our AHC experiments. Beyond a few visible outliers, the sample by feature heatmap shows largely correlated patterns across many samples, suggesting that HOG does not provide sufficiently discriminative structure for separating the nail conditions. This is consistent with the quantitative results in Table 5.17 and the visual evidence in Fig. 5.5.

Tabel 5.17: AHC results using HOG features on the training dataset.

n_clusters	linkage	metric	accuracy	accuracy	binary	ARI	NMI
4	ward	euclidean	0.28		0.98	0.05	0.14
4	average	euclidean	0.21		0.91	-0.00	0.00
4	average	manhattan	0.28		0.98	0.05	0.13
4	complete	euclidean	0.20		0.91	-0.00	0.02
4	complete	manhattan	0.21		0.91	0.01	0.03
6	ward	euclidean	0.27		0.98	0.05	0.13
6	average	euclidean	0.21		0.91	-0.00	0.00
6	average	manhattan	0.28		0.98	0.05	0.12
6	complete	euclidean	0.26		0.91	0.03	0.05
6	complete	manhattan	0.21		0.91	0.01	0.03
8	ward	euclidean	0.27		0.98	0.05	0.12
8	average	euclidean	0.20		0.91	-0.00	0.01
8	average	manhattan	0.25		0.98	0.05	0.11
8	complete	euclidean	0.23		0.91	0.03	0.05
8	complete	manhattan	0.21		0.91	0.02	0.05
10	ward	euclidean	0.24		0.98	0.05	0.11
10	average	euclidean	0.20		0.91	-0.00	0.01
10	average	manhattan	0.22		0.98	0.05	0.11
10	complete	euclidean	0.22		0.91	0.03	0.05
10	complete	manhattan	0.19		0.91	0.02	0.05

Tabel 5.18: AHC results using HOG features on the validation dataset.

n clusters	linkage	metric	accuracy	accuracy	binary	ARI	NMI
6	WARD	EUCLIDEAN	0.45		0.96	0.18	0.31
4	AVERAGE	MANHATTAN	0.44		0.88	0.20	0.24
6	AVERAGE	MANHATTAN	0.44		0.88	0.21	0.28
4	WARD	EUCLIDEAN	0.42		0.96	0.19	0.30
8	WARD	EUCLIDEAN	0.43		0.96	0.22	0.32

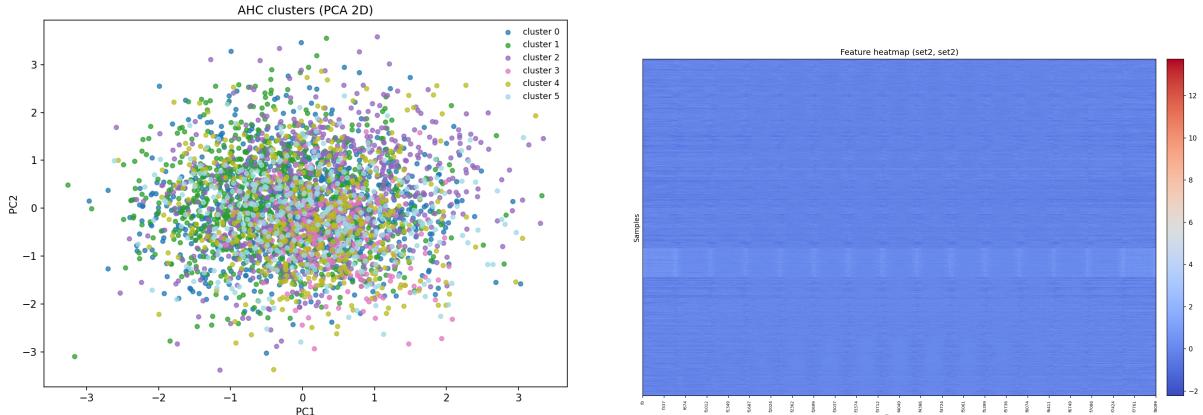


Figure 5.5: AHC results with HOG features summarized in Table 5.17.

5.4 SVM

5.4.1 HSV

Tabel 5.19: *SVM validation results for HSV features.*

kernel	C	degree	coef0	accuracy	f1
LINEAR	0.10	–	–	0.45	0.44
LINEAR	1	–	–	0.40	0.37
LINEAR	10	–	–	0.26	0.25
RBF	0.10	–	–	0.29	0.24
RBF	0.10	–	–	0.27	0.23
RBF	1	–	–	0.47	0.45
RBF	1	–	–	0.52	0.50
RBF	10	–	–	0.49	0.47
RBF	10	–	–	0.52	0.49
POLY	0.10	2	0.00	0.19	0.08
POLY	0.10	3	0.00	0.18	0.06
POLY	1	2	1.00	0.49	0.48
POLY	1	3	1.00	0.48	0.47
POLY	10	2	0.00	0.43	0.41
POLY	10	3	0.00	0.37	0.38
POLY	10	2	1.00	0.55	0.53

5.4.2 LBP

Tabel 5.20: *SVM validation results for HSV features.*

kernel	C	degree	coef0	accuracy	f1
LINEAR	0.10	–	–	0.54	0.51
LINEAR	1.00	–	–	0.62	0.59
LINEAR	10.00	–	–	0.55	0.52
RBF	0.10	–	–	0.54	0.45
RBF	10.00	–	–	0.65	0.60
RBF	10.00	–	–	0.65	0.60
POLY	0.10	2	0.00	0.48	0.36
POLY	0.10	3	0.00	0.44	0.41
POLY	0.10	2	1.00	0.47	0.45
POLY	1.00	2	0.00	0.52	0.43
POLY	10.00	2	1.00	0.55	0.51
POLY	10.00	3	1.00	0.60	0.55

Chapter 6

Conclusion

Across our experiments on the Nails dataset, we observed clear differences in how well each handcrafted representation supports clustering. Overall, LBP were the most useful, particularly for distinguishing *healthy* from *unhealthy* samples. In contrast, the multiclass setting proved harder to improve, disease categories often overlap, making it difficult to obtain clusters that align cleanly with the six ground truth labels. HSV histograms ranked second, their performance appears limited by background artifacts (padding and surrounding skin), it is possible that a tighter nail only segmentation could further improve color based separability. Finally, HOG features were the least effective in our experiments, likely because gradient based descriptors capture global edge structure that is not sufficiently distinctive.

Regarding models, Agglomerative Hierarchical Clustering (AHC) produced the most interpretable multiclass results, resulting in structured partitions that were easier to analyze than DBSCAN, which frequently assigned many points to noise. For the binary task, the most successful configuration combined LBP features with clustering, enabling a relatively clear separation of healthy nails, the healthy class could be mapped with approximately 75-80% accuracy, which was better then the random guess.

Compared with the SVM, we seen weaker results using the unsupervised methods, given that we can linearly separate the classes with a high accuracy.

As future work, it would have been interesting to explore t-SNE, which may better preserve structure of our datapoints.