
Supplementary Material for DermaCon-IN

Shanawaj S Madarkar* **Mahajabeen Madarkar*** **Madhumitha V*** Teli Prakash
Konda Reddy Mopuri Vinaykumar MV KVL Sathwika Adarsh Kasturi
Gandla Dilip Raj PVN Supranitha Harsh Udai

Contents

A Dataset Details	2
A.1 Rook’s Classification for a Hierarchical Framework	2
A.2 Descriptor Design and Clinical Alignment	2
A.3 Choice of Anatomical Site	4
A.4 Comparative Coverage of Dermatological Disease Categories Across Public Datasets	5
A.5 Other stats of the DermaCon-IN dataset	5
B Model Architecture Details	6
B.1 Concept Bottleneck Modeling	6
B.1.1 1-level CBM for MC/ SC Classification	6
B.1.2 2-level CBM for MC & SC Classification	7
B.1.3 Handling Variable Image Resolutions.	9
B.2 Detailed Performance Analysis	9
C More examples for Post-hoc analysis	11
C.1 Concept Activation Analysis	11
C.2 Concept Contribution Analysis Across Semantically Disjoint Families	11
C.3 Concept-specific Spatial Attribution via Grad-CAM	14
D Cross-Dataset Validation and Distributional Coverage	16
E Ethical Considerations	16
F Societal Impact	17
G Glossary of Terms	17

*Equal contribution

A Dataset Details

A.1 Rook’s Classification for a Hierarchical Framework

Most dermatology datasets, such as *SD-198* [33] and *SCIN* [40], employ flat diagnostic label structures without an overarching clinical taxonomy. Others, like *Fitzpatrick17k* [10] and *SkinCon* [5], offer fine-grained or concept-level annotations but do not embed these within an etiologically structured or pathophysiology-aware hierarchy. As a result, they fall short of modeling the layered reasoning typical of clinical diagnosis. While sufficient for benchmarking classification models, these taxonomies lack clinical depth and fail to represent the structured reasoning used in dermatological diagnosis.

Clinical Relevance: Rook’s taxonomy [9] introduces a hierarchical framework grounded in disease etiology and pathological processes, organizing conditions into superclasses such as *infectious*, *inflammatory*, *pigmentary*, and *appendageal*, etc. This mirrors how clinicians formulate differential diagnoses from broad mechanisms to specific entities and enables AI systems to produce outputs that resonate with clinical logic.

Contextual Suitability: In Indian outpatient dermatology, where fungal, pigmentary, and inflammatory conditions predominate, this structure offers strong alignment with real-world case loads. Unlike neoplasm-focused Western datasets, Rook’s [9] schema reflects the diversity and prevalence of diseases encountered in Low & Middle Income Countries (LMIC) settings, while remaining extensible to global contexts.

Curation and Learning Advantages: A hierarchical setup facilitates consistent labeling, balanced representation across superclasses, and structured sampling strategies. It also supports coarse-to-fine prediction models, multi-task learning, and scalable dataset expansion.

Interoperability with Clinical Systems: Crucially, Rook’s classification aligns with international standards such as the WHO’s ICD-10 [41] and ICD-11, which also structure diseases by etiology and anatomical relevance. Since ICD codes underpin electronic health records, billing, and decision support tools, adopting a Rook-aligned hierarchy ensures that the dataset remains interoperable with clinical infrastructures. This enables seamless integration into healthcare workflows, extending its utility well beyond academic modeling to real-world deployment.

A.2 Descriptor Design and Clinical Alignment

Descriptor Design and Selection: We annotate 5,450 dermatological images using 47 carefully selected descriptors encompassing primary lesions, secondary changes, pigmentary alterations, vascular anomalies, and surface/nail abnormalities. The taxonomy is rooted in established dermatological frameworks [3], ensuring clinical interpretability and consistency with medical curricula and diagnostic practice. While many descriptors overlap with those used in *SkinCon* [5] (e.g., *papule*, *plaque*, *vesicle*), our set incorporates necessary refinements based on regional prevalence and diagnostic relevance. Descriptors with high clinical utility but low frequency (e.g., *poikiloderma*, *gray*, *salmon*) were retained to preserve diagnostic granularity. The descriptors such as *translucent*, *friable*, and *dome-shaped* lacked distinct visual features in 2D clinical images and so were not included in the list of descriptors, which were otherwise part of *SkinCon* descriptors.

Contextual Tailoring for Indian Dermatology: Our descriptor set is designed with sensitivity to Indian outpatient dermatology. Conditions such as vitiligo and post-inflammatory hyperpigmentation necessitate precise pigmentary annotations (*white*, *gray*, *brown*, *pigmented*). Similarly, endemic infections (e.g., scabies, impetigo) justify the inclusion of *crust*, *burrow*, and *abscess*. Chronic inflammatory presentations (*lichenification*, *hyperkeratotic plaques*) and nail findings (*pitted nail*, *discolored nail*) are also integrated based on their diagnostic frequency and clinical relevance in South Asia.

Standardization and Future Compatibility: Despite its regional grounding, our descriptor vocabulary remains aligned with global dermatological standards, enabling seamless integration with future datasets developed under similar clinically guided frameworks. By leveraging literature-backed

terminology, our schema offers both backward compatibility with existing datasets like SkinCon [5] and forward compatibility with evolving multimodal dermatology benchmarks.

Summary. Our descriptor design balances regional specificity with standardization. It enhances diagnostic interpretability, improves label quality for model training, and facilitates interoperability across future dermatological datasets. A detailed table of descriptors, definitions, and rationale is provided in Sup. Table 1.

Table 1: Complete Concept Descriptor Table with Definitions and Reasoning.

S.No	Concept	Explanation	Reason for Choosing the Descriptor
1	Abscess	A localized collection of pus within the dermis or subcutaneous tissue, typically surrounded by inflamed tissue; clinically presents as a painful, erythematous, fluctuant nodule.	Common in bacterial infections; present frequently in tropical and humid climates.
2	Acuminate	A lesion with a tapering, pointed shape; commonly seen in viral warts.	Seen in viral warts and common sexually transmitted infections like condyloma in Indian OPDs.
3	Atrophy	A reduction or thinning of tissue, which may involve the epidermal (outer skin layer), dermal (middle layer), or subcutaneous (deep fat and connective tissue layer).	Frequently observed in chronic corticosteroid use and dermatoses like lichen sclerosus.
4	Brown (Hyperpigmentation)	Darkened area of the skin due to excess melanin.	Extremely common in post-inflammatory states and melasma in the Indian population.
5	Bulla	A circumscribed lesion > 1 cm in diameter that contains liquid (clear, serous or haemorrhagic).	Seen in autoimmune and infective blistering disorders; important in differential diagnosis.
6	Burrow	It is a thin, wavy, slightly raised line on the skin, typically found between the fingers or toes, caused by the scabies mite burrowing under the surface.	Highly relevant in scabies, which has a significant endemic prevalence in India.
7	Comedo	Comedo is a blocked skin pore caused by the accumulation of oil and dead skin cells. An open comedo, commonly known as a blackhead, has an exposed surface where the trapped material darkens due to air exposure. A closed comedo, or whitehead, remains sealed under the skin, forming a small bump.	Key feature in acne, a major dermatological complaint in adolescents and young adults.
8	Crust	Dried serum, blood or pus on the surface of the skin.	Present in infected eczemas and impetigo, common in Indian children.
9	Cyst	Cyst is a papule or nodule that contains fluid or semi-fluid material, making it soft and fluctuant to touch.	Frequent presentation in both cosmetic and inflammatory conditions, like epidermoid cysts.
10	Dilated Vein	Enlarged, visible superficial vein; may be seen in varicosities or venous insufficiency.	Observed in varicose conditions and vascular anomalies, especially in rural patients.
11	Discolored Nail	Alteration in nail color due to trauma, infection, pigmentation, or systemic disease.	an Indicator of systemic illness, fungal infections, or trauma-related disorders.
12	Edema	Swelling of the skin due to fluid accumulation in the dermis or subcutaneous tissue.	Secondary signs in infections, inflammatory conditions, and systemic diseases.
13	Erosion	Superficial loss of the epidermis, often due to ruptured blisters; heals without scarring.	Occurs in healing blisters and vesiculobullous diseases seen in Indian settings.
14	Erythema	Redness of the skin due to increased blood supply.	Fundamental indicator of inflammation; often subtle in pigmented skin.
15	Excoriation	A loss of the epidermis and a portion of the dermis due to scratching or an exogenous injury.	Seen in pruritic disorders like scabies, atopic dermatitis, and lichen simplex chronicus.
16	Exophytic/Fungating	It describes a type of lesion that grows rapidly, breaks through the skin surface, and often appears ulcerated, foul-smelling, and infected, resembling a fungus-like mass.	Helps classify malignant and advanced skin lesions; relevant in tertiary care.
17	Exudate	Oozing fluid composed of serum, pus, or blood; typically due to inflammation or infection.	Observed in infected wounds, ulcers, and pyodermas.
18	Fissure	Fissure is a linear crack or break in the skin that begins in the outermost layer (stratum corneum) and may extend into the deeper dermis, often causing pain or bleeding.	Very common in xerotic skin conditions and hand/foot eczema exacerbated by occupational exposure.
19	Flat-topped	Lesion with a flattened horizontal surface; characteristic of lichen planus.	Key for diagnosing lichen planus, prevalent in Indian adults.
20	Gray	A color descriptor typically indicating post-inflammatory hyperpigmentation or pigment incontinence.	Represents pigment incontinence or deeper melanin deposition; seen in pigmented disorders.
21	Hair Patch	Hair patch refers to a localized area on the skin where hair is either abnormally present (increased density or unusual location) or absent (loss of hair in a defined area)	Helps identify abnormal loss or growth of hair in a specific area.
22	Hyperkeratotic plaques	Thickened plaques with an excessive build-up of keratin.	Observed in psoriasis, lichen simplex, and chronic eczema, common in Indian clinics.

Continued on next page

Table 1: Complete Concept Descriptor Table (continued)

S.No	Concept	Explanation	Reason for Choosing the Descriptor
23	Induration	It refers to an area of the skin or tissue that feels firm or hardened to the touch, without any underlying calcification or bone formation.	Helps differentiate infections (e.g., cellulitis) or granulomatous conditions (e.g., leprosy).
24	Lichenification	It is thickened, rough skin with exaggerated skin lines, typically resulting from repeated rubbing or scratching	Common in chronic atopic and lichen simplex; consequence of habitual scratching.
25	Macule	A flat, circumscribed, nonpalpable lesion that differs in colour from the surrounding skin.	Key to diagnosing pigmentary and vascular disorders like vitiligo and leprosy.
26	Nodule	An elevated, solid, palpable lesion > 1 cm usually located primarily in the dermis and/or subcutis.	Important for deep fungal infections, cutaneous TB, or cystic swellings in endemic areas.
27	Papule	An elevated, solid, palpable lesion that is ≤ 1 cm in diameter.	Seen in common dermatoses like folliculitis, acne, and viral warts.
28	Patch	A large area of colour change, with a smooth surface.	Central to identifying vitiligo, pityriasis alba, and leprosy.
29	Pedunculated	Lesion attached by a narrow stalk.	Helps in the classification of benign tumors like acrochordons or neurofibromas.
30	Pigmented	Lesions exhibiting increased pigment; may be brown, gray, or black.	Essential in differentiating dermatoses on brown skin; pigmentary presentations dominate.
31	Pitted Nail	Small depressions on the nail surface.	Common signs of nail psoriasis and alopecia areata.
32	Plaque	A circumscribed, palpable lesion ≥ 1 cm in diameter; most plaques are elevated.	Describes major lesion morphology in tinea, psoriasis, and lichen simplex.
33	Poikiloderma	Simultaneous presence of atrophy, telangiectasia and hypo and hyperpigmentation.	Seen in late-stage connective tissue diseases; needs documentation in atypical Indian cases.
34	Purpura/Petechiae	Haemorrhage into the skin due to pathological processes, primarily of blood vessels.	Observed in vasculitis and hematological disorders.
35	Pustule	A circumscribed lesion that contains pus.	Seen in acne, folliculitis, and impetigo; frequently present in outpatient cases.
36	Salmon	Pink-orange hue used to describe psoriatic lesions, especially on lighter skin tones.	Color reference for certain psoriatic plaques in lighter Indian skin tones.
37	Scale	A visible accumulation of keratin, forming a flat plate or flake.	Typical in dermatophytosis, psoriasis, and seborrheic dermatitis; common in humid climates.
38	Scar	Fibrotic replacement of normal skin architecture after injury.	Important to track disease healing and secondary changes post-injury or intervention.
39	Striae	Linear atrophic lesions due to dermal tearing.	Common due to corticosteroid use, obesity, puberty, and pregnancy-related changes.
40	Telangiectasia	Permanently dilated capillaries.	Seen in rosacea, lupus, and long-term corticosteroid use.
41	Ulcer	Full-thickness loss of the epidermis plus at least a portion of the dermis.	Crucial for identifying diabetic foot, leprosy, and chronic venous ulcers.
42	Vesicle	A circumscribed lesion ≤ 1 cm in diameter that contains liquid (clear, serous or haemorrhagic).	Key feature in varicella, dermatitis herpetiformis, and contact dermatitis.
43	Warty	Verrucous surface resembling a wart; rough and irregular.	Descriptive of HPV-induced lesions and seborrheic keratoses, which are common in the elderly in India.
44	Wheal	A transient elevation of the skin due to dermal edema.	Seen in urticaria due to infections, drugs, and food reactions.
45	White (Hypopigmentation)	Lighter than normal skin color due to loss or reduction of melanin.	Central to vitiligo, pityriasis alba, and tinea versicolor, which are frequent in India.
46	Xerosis	Abnormal dryness of the skin.	Highly prevalent due to hygiene practices, hard water, and low humidity in winter.
47	Yellow	Describes lesions with lipid, keratin, or bile pigment.	Seen in xanthomas, sebaceous discharge, and bacterial pustules.

A.3 Choice of Anatomical Site

We developed our body region taxonomy by aligning it with how dermatologists reason through diagnoses in clinical settings. Standard references such as *Rook’s Dermatology* [9], *Fitzpatrick’s Dermatology*, and *Bolognia’s Dermatology* [3] consistently describe skin conditions based on their anatomical distribution. These texts emphasize that lesion location plays a central role in diagnosis, distinguishing, for example, mucosal from cutaneous presentations. By mirroring these clinically grounded patterns, we ensured that our descriptors reflect the spatial logic used in real-world diagnostic workflows.

Coverage gaps in existing datasets: We reviewed widely used dermatology datasets—ISIC, Fitzpatrick17k [10], SD-198 [33], etc., and found that they often lack precise anatomical context. Most provide cropped images that obscure lesion location or annotate broad categories like “face” or “limb,” limiting their clinical utility. To address this, we explicitly included underrepresented yet diagnostically critical regions such as the armpits and groin area. These regions are key to diagnosing conditions like candidiasis and tinea infections.

Hierarchical design for diagnostic reasoning: We designed our annotation hierarchy to support models that reason at multiple anatomical resolutions. Dermatologists often start with coarse region-based hypotheses and further refine them until morphology and context become clearer. Our taxonomy enables similar flexibility, allowing models to learn general patterns at macro levels while capturing fine-grained distinctions when needed. This structure mirrors how clinicians disambiguate conditions with overlapping visual features based on location. The complete list of anatomical regions used in our annotation schema is provided in the Datasheet included with the supplementary material for reference.

A.4 Comparative Coverage of Dermatological Disease Categories Across Public Datasets

Most publicly available dermatology datasets were developed for specific diagnostic tasks, often skin cancer triage, and do not reflect the full diagnostic spectrum seen in routine outpatient clinics, particularly in low and middle-income countries (LMICs) where poor maintenance of hygiene amongst the population is a key factor for the spread of infectious diseases. As a result, critical categories such as infectious disorders, pigmentary changes, and appendageal conditions are underrepresented or absent entirely. To assess where our dataset stands in relation to existing benchmarks, we compiled an exhaustive comparison in the Sup. Table 2 of major dermatology datasets across clinically meaningful diagnostic categories.

Table 2: Comparative distribution (in %) of dermatological disease categories across public datasets. Only our dataset provides a good representation across all categories as encountered in routine outpatient care, including mixed diagnoses.

Dataset	Infectious	Inflammatory (incl. keratinisation)	Pigmentary	Appendageal (Acne/Hair)	Neoplastic	No Definite Diagnosis	Others	Key Observations
DermaCon-IN	40.86	30.53	15.25	8.81	0.95	0.68	2.92	Full-spectrum, OPD-aligned dataset designed for LMIC clinical diversity
SCIN	21.85	56.5	0.75	Sparse	5.20	Not included	15.0	Inflammatory-heavy; lacks uncertain and appendageal representation
Fitzpatrick17k	–	65.67	7.2	Sparse	26.7	Not included	–	Inflammatory and neoplastic skew; lacks diagnostic uncertainty
PASSION	63.52	25.05	Sparse	–	–	Not included	11.43	Pediatric LMIC dataset; lacks neoplastic and appendageal coverage
DDI	1.66	2.28	0.46	Sparse	92.7	Not included	3.0	Biopsy-focused dataset; neoplasm-dominant
ISIC 2020	–	–	–	–	100.00	Not included	–	Exclusive neoplasm dataset
PH2	–	–	–	–	100.00	Not included	–	Exclusive neoplasm dataset
HAM10k	–	–	–	–	100.00	Not included	–	Exclusive neoplasm dataset
PAD-UFES	–	–	–	–	100.00	Not included	–	Cancer-focused dataset only

A.5 Other stats of the DermaCon-IN dataset

Demographic and Phenotypic Distribution: The dataset reveals distinct demographic and phenotypic trends across age, sex, and skin tone that shape its clinical composition (Sup. Figure 1). Most images are concentrated in the 20–40 age group, with secondary peaks in 10–20 and 40–60, reflecting outpatient demand among working-age individuals. Children (0–10) and older adults (60+) are relatively underrepresented. A notable sex imbalance exists, with males contributing more samples (3394 vs. 2056), possibly due to sociocultural factors. Infectious and pigmentary disorders dominate across both sexes, with females also showing higher counts for appendageal and inflammatory conditions.

In terms of skin tone, Fitzpatrick Types 4 and 5 are most prevalent, followed by Type 3. Type 6 is modestly represented (136 images), but is still less frequent relative to mid-tone categories. The Monk Skin Tone distribution similarly centers on MST 6 and 7, with lighter (MST 4) and darker tones (MST 8–9) less represented. Across all groups, infectious, inflammatory, and pigmentary disorders remain most common.

Concept Distribution: A detailed analysis of the concept counts per Main class is presented in the Sup. Fig. 2. The figure demonstrates the occurrence of shared concepts across the main class. Further, it also suggests an imbalance of concepts within each class.

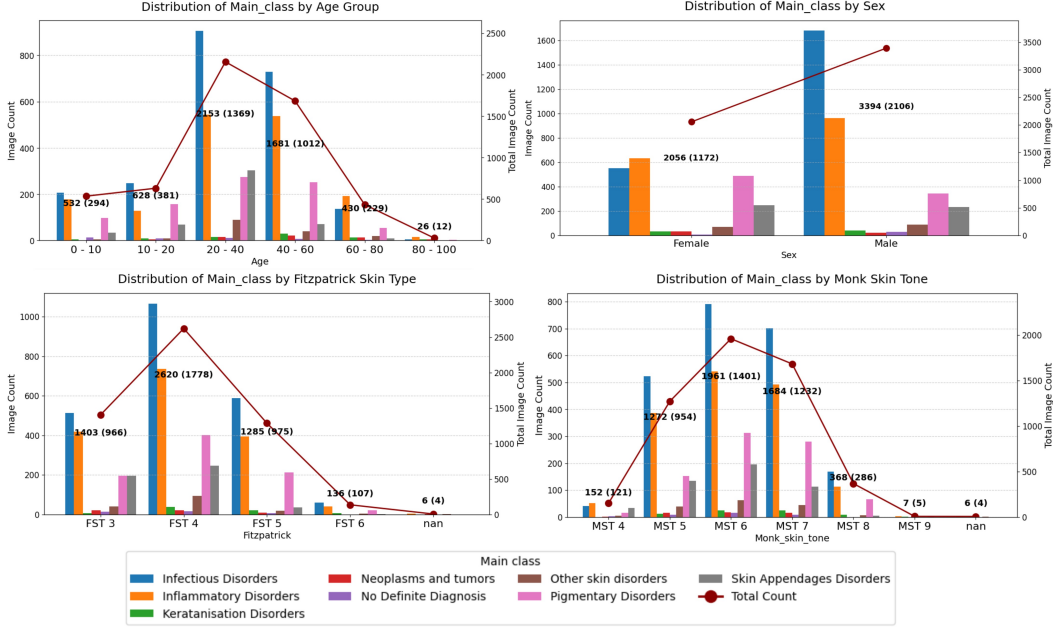


Figure 1: Plots for distribution of dermatological Main Classes (MC) across patient demographics. Each bar represents the number of images per category (Age group, Fitzpatrick type, Monk Skin Tone, or Sex), while the red line denotes the total number of images. Numeric annotations above each point indicate image count followed by subject count in parentheses.

B Model Architecture Details

B.1 Concept Bottleneck Modeling

We investigate two modeling variants under the Concept Bottleneck framework: a 1-level CBM that performs either main class (MC) or sub-class (SC) prediction independently, and a 2-level CBM that jointly predicts both levels in a hierarchical structure. While both designs share a common concept encoder comprising a Swin Transformer followed by dropout and a linear projection to produce concept logits c^l , they differ in their classification strategies as discussed below. Architectural and training hyperparameters for all CBM variants are summarized in Sup. Table 3.

B.1.1 1-level CBM for MC/SC Classification

To explore the utility of medically grounded concepts in isolation, we train a Concept Bottleneck Model (CBM) that predicts dermatological categories, either main classes (MC) or sub-classes (SC), directly from concept representations. This model maps image features to concept logits and performs classification using those logits alone.

Model Architecture. An input image $x \in \mathbb{R}^{3 \times H \times W}$ is first passed through a Swin Transformer backbone to extract a feature embedding. This embedding is then followed by a fully connected projection having Dropout, to produce a concept logit vector:

$$c^l = \text{Linear}(\text{Dropout}(\text{SwinTransformer}(x))) \in \mathbb{R}^K$$

The raw concept logits c^l are used as input to a linear classification head:

$$y = f(c^l)$$

where $f(\cdot)$ is a single-layer classifier mapping to either subclass or main class logits depending on the training objective.

Training Objective. The model is trained with two primary losses:

$$\begin{aligned} \mathcal{L}_{\text{concept}} &= \text{BCEWithLogits}(c^l, c_{\text{true}}) \\ \mathcal{L}_{\text{class}} &= \text{CrossEntropy}(f(c^l), y_{\text{true}}) \end{aligned}$$

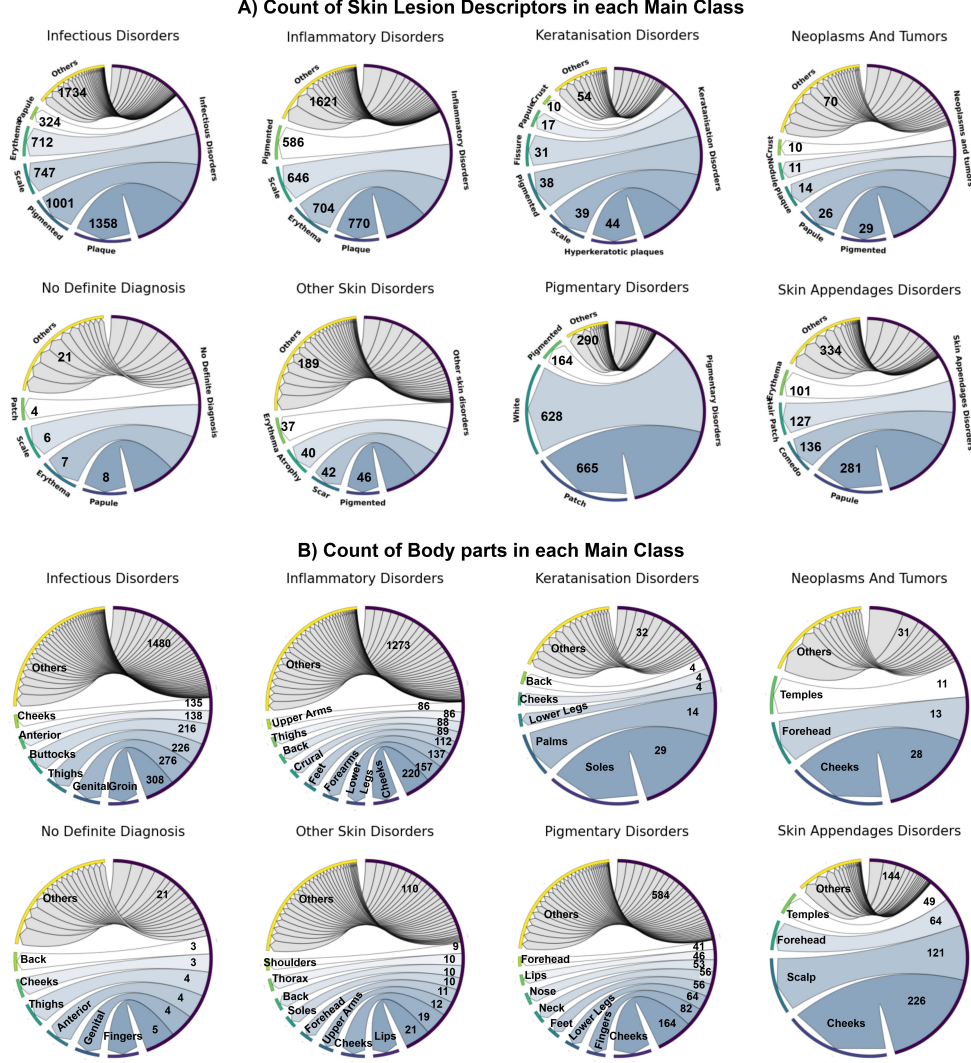


Figure 2: Circular plots of counts of Concepts (both skin lesion descriptors and body parts) for each main class skin labels.

The total loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{concept}} + \mathcal{L}_{\text{class}} + \lambda \|c^l\|_1$$

An L1 regularization term encourages sparsity in the concept representation, invoking the relevant concepts, improving interpretability, and acting as a regularizer during optimization.

Sampling and Imbalance Handling. To account for label imbalance in the main class distribution, we compute per-sample weights based on the inverse frequency of the main class and use a WeightedRandomSampler during training.

B.1.2 2-level CBM for MC & SC Classification

We propose a two-level hierarchical prediction framework based on Concept Bottleneck Models (CBMs), where disease predictions are structured via an interpretable intermediate layer of clinical concepts (refer Fig. 7 of the main paper). The task is to jointly predict a fine-grained dermatological *sub-class* label $y^1 \in \{1, \dots, N_{\text{sub}}\}$ and a coarser *main class* label $y^2 \in \{1, \dots, N_{\text{main}}\}$, from an input image $x \in \mathbb{R}^{3 \times H \times W}$.

Table 3: Training and architecture hyperparameters for all CBM models.

Component	Value / Setting
Backbone architecture	Swin-B (patch4, window12, 384)
Image resolution	512×512
Concept vector dimension K	96
Dropout before concept head	0.3
Concept supervision loss	BCEWithLogits (sigmoid applied internally)
Task heads $f(\cdot), g(\cdot)$	Linear layers (no hidden layer)
SC/MC loss	CrossEntropy with label smoothing $\epsilon = 0.1$
L1 regularization weight λ	1×10^{-4}
Optimizer	AdamW
Learning rate	3×10^{-5}
Weight decay	1×10^{-4}
Scheduler	CosineAnnealingLR ($T_{\max}=30$, min LR= 10^{-6})
Training epochs	40
Batch size	130 (2-level), 100 (1-level)
Multi-GPU support	DataParallel (4 GPUs)

Concept Space. A vision backbone (Swin Transformer) maps the input image to a K -dimensional concept logit vector:

$$c^l = \text{Linear}(\text{Dropout}(\text{SwinTransformer}(x))) \in \mathbb{R}^K$$

We define $c = \sigma(c^l) \in [0, 1]^K$ as the sigmoid-activated concept probabilities, which are supervised using binary multi-label concept annotations. The raw logits c^l , not the sigmoid outputs, are further used for downstream disease classification, preserving gradient flow and avoiding saturation effects.

Type 1: Cascade Architecture. This model enforces taxonomy consistency by chaining predictions through intermediate subclass logits:

$$y^1 = f(c^l), \quad y^2 = g(y^1)$$

Here, $f(\cdot)$ is a linear layer mapping concept logits to subclass logits $\in \mathbb{R}^{N_{\text{sub}}}$, and $g(\cdot)$ is another linear layer mapping subclass logits to main class logits $\in \mathbb{R}^{N_{\text{main}}}$. This design structurally enforces taxonomic consistency between sub- and main classes, aligning with medical hierarchies. However, performance is inherently constrained by the reliability of the first-stage prediction y^1 , making it prone to error propagation.

Type 2: Parallel Architecture. Instead of sequential dependency, both SC and MC are predicted directly from concept logits:

$$y^1 = f(c^l), \quad y^2 = g(c^l)$$

Here, $f(\cdot)$ and $g(\cdot)$ are task-specific linear classifiers. This decouples the learning paths while maintaining shared semantic grounding via concepts. The architecture benefits from multi-task supervision and avoids dependency on intermediate task outputs. It allows the model to flexibly learn patterns that are specific to either task while still being grounded in a common, interpretable representation.

Loss Function. The overall objective for each image includes:

- Binary cross-entropy loss on sigmoid-transformed concepts:

$$\mathcal{L}_{\text{concept}} = \text{BCEWithLogits}(c^l, c_{\text{true}})$$

- Cross-entropy loss on subclass logits:

$$\mathcal{L}_{\text{SC}} = \text{CrossEntropy}(f(c^l), y_{\text{true}}^1)$$

- Cross-entropy loss on main class logits:

$$\mathcal{L}_{\text{MC}} = \begin{cases} \text{CrossEntropy}(g(y^1), y_{\text{true}}^2), & \text{Type 1} \\ \text{CrossEntropy}(g(c^l), y_{\text{true}}^2), & \text{Type 2} \end{cases}$$

- L1 regularization on concept logits to promote sparsity:

$$\mathcal{L}_{L1} = \lambda \|c^l\|_1$$

The total loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{concept}} + \mathcal{L}_{\text{SC}} + \mathcal{L}_{\text{MC}} + \mathcal{L}_{L1}$$

Class Imbalance Handling. To account for imbalanced subclass and main class distributions, we assign sample-wise weights based on inverse frequency. For each sample i , the final weight is computed as:

$$w_i = \frac{1}{2} \left(\frac{1}{f(y_i^1)} + \frac{1}{f(y_i^2)} \right)$$

where $f(\cdot)$ is the empirical class frequency. These weights are used with a `WeightedRandomSampler` to ensure that rare classes contribute adequately during training.

B.1.3 Handling Variable Image Resolutions.

Dermatological images in real-world clinical datasets often exhibit heterogeneous resolutions and aspect ratios. To address this, we first resize images such that their longer side does not exceed 512 pixels while maintaining the aspect ratio. The resized image is then zero-padded to 512×512 to ensure uniform input dimensions across batches. This approach preserves image content without distortion and enables efficient batch processing while maintaining compatibility with the fixed input size expected by the Swin Transformer backbone. Padding is handled dynamically during training and validation through a custom collate function.

B.2 Detailed Performance Analysis

Main Class Classification Performance. Sup. Table 4 presents a detailed evaluation of our best-performing models for main class (MC) prediction. We report Top-1, Top-3, and Top-5 classification accuracies along with macro AUC values (%) across all main classes and per-class breakdowns.

For example, in inflammatory and pigmentary disorders, Top-1 accuracies range between 65–73%, but Top-5 scores consistently rise above 95%, indicating that the correct class is almost always among the top few ranked predictions. This pattern holds across infectious and appendageal disorders as well, with Top-5 accuracy approaching ceiling levels (above 97% in most cases). Such trends affirm the model’s capacity to capture relevant features even in challenging differential diagnosis.

Conversely, classes such as Keratinization disorders, Neoplasms, others, and no diagnosis exhibit lower Top-1 performance, yet still benefit from 20–40% absolute improvement when considering Top-5 predictions. This suggests that while these categories are harder to pinpoint as the top choice, the model often recognizes them as plausible alternatives, supporting their inclusion in a ranked differential. This reflects the inherent difficulty or class imbalance in these categories. These results highlight strong discriminative performance in clinically prevalent and visually distinct categories, while also pointing to the need for improvement in more ambiguous or underrepresented classes.

Table 4: Detailed performance analysis of best models for main class (MC) prediction. All metrics include Accuracy (Top-1,3,5) and AUC (ovr) as % for both: all classes and per class.

Model	Metric	All classes	Infectious	Inflammatory	Keratanisation	Neoplasms	No Diagnosis	Other	Pigmentary	Appendages
Swin (MC)	Top-1	70.88	77.51	68.51	14.29	16.67	0.00	14.29	73.17	81.19
	Top-3	91.06	94.98	96.10	50.00	83.33	33.33	32.14	87.20	92.08
	Top-5	94.86	97.13	98.70	50.00	83.33	66.67	60.71	92.07	97.03
	AUC	78.36	84.80	82.76	66.29	87.05	55.55	63.57	91.11	95.79
CBM (Concepts+ MC)	Top-1	67.55	72.49	67.21	7.14	25.00	0.00	7.14	70.73	77.23
	Top-3	91.72	98.33	96.75	21.43	41.67	16.67	32.14	89.02	90.10
	Top-5	96.57	99.52	99.68	57.14	83.33	66.67	60.71	94.51	97.03
	AUC	79.17	85.70	81.80	56.40	79.10	74.34	68.60	93.09	94.37
CBM Type 2 (Concepts+ MC & SC)	Top-1	70.12	79.43	66.23	7.14	16.67	0.00	14.29	71.95	75.25
	Top-3	90.49	94.50	96.43	14.29	41.67	33.33	46.43	87.80	92.08
	Top-5	95.53	95.93	99.03	64.29	75.00	66.67	82.14	95.73	95.05
	AUC	78.71	84.78	83.08	66.30	80.62	60.91	68.39	91.77	93.82

Multi-disease Co-occurrence Analysis. A small but clinically meaningful subset contains concurrent lesions from more than one disease type on the same anatomical site (e.g., *Inflammatory + Fungal*, *Fungal + Bacterial*). We encode these as dedicated second-level subclasses to let models explicitly learn multi-disease patterns and disentangle overlapping cues. To assess behavior on these cases, we performed a targeted misclassification analysis with the best *Swin Transformer*, recording whether predictions matched both constituent types (*Predicted as Both*), only one (*Predicted as Either*), or neither (*Predicted as Other*). Results are summarized in Table 5.

Table 5: Misclassification analysis for multi-disease co-occurrence samples. “Predicted as Both” denotes a correct assignment to the multi-disease subclass; “Predicted as Either” lists counts that match one constituent type; “Predicted as Other” lists counts for unrelated classes.

True Class	Predicted as Both	Predicted as Either	Predicted as Other
Inflammatory + Infectious– Bacterial	5	Inflammatory (4), Bacterial (3)	Pigmentary (1)
Fungal + Bacterial	1	Fungal (3), Inflammatory (1)	—
Parasitic + Bacterial	3	Parasitic (2), Inflammatory (1)	—
Inflammatory + Fungal	—	Fungal (2), Inflammatory (1)	—

Overall, a subset of samples is correctly recognized as their multi-disease subclass (*Predicted as Both*). Many are assigned to one constituent type (*Predicted as Either*), likely reflecting dominance of one pathology’s visual cues (e.g., markedly scaly plaques in fungal disease). A smaller number maps to unrelated categories (*Predicted as Other*). These findings underscore the dataset’s clinical realism and motivate **multi-label, context-aware** approaches to robustly handle co-occurring dermatological conditions.

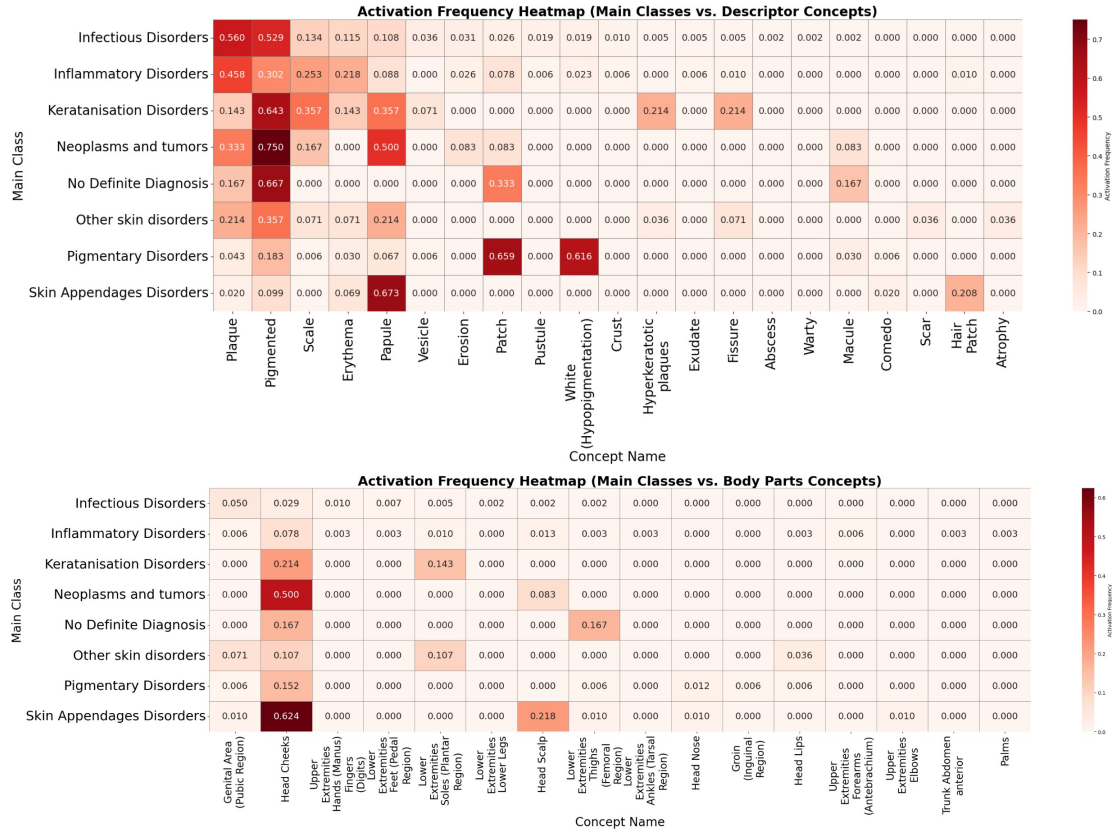


Figure 3: Concept frequency heatmaps for the analysis of active concepts per class (Main class) on the full test set.

C More examples for Post-hoc analysis

The following post-hoc analyses are conducted using the best-performing CBM, namely the Type-2 model, on the test set provided in GitHub.

C.1 Concept Activation Analysis

We analyzed the activation patterns of interpretable concepts across eight dermatological disorder categories to evaluate semantic alignment in model reasoning. Concepts were grouped into two clinically motivated families: *descriptor* concepts (e.g., *plaque*, *erythema*, *vesicle*) and *body part* concepts (e.g., *head*, *cheek*, *extremities*). For each class, activation frequency was computed as the fraction of samples per class in which a given concept was predicted to be positively active and is presented as heatmaps in Sup. Fig. 3. Concepts with negative activation across all classes were excluded from visualization.

Out of 47 available descriptor concepts, 21 exhibited positive activation. For body part concepts, 16 out of 49 exhibited positive activation. This indicates that the model’s decision process selectively emphasizes a small subset of clinically relevant features while disregarding many others, which could be due to lower predictive value.

Notably, descriptor activations showed meaningful alignment with known disease characteristics: *papule* and *pigmented* were strongly associated with *Skin Appendages Disorders* and *Neoplasms and Tumors* respectively, while *Pigmentary Disorders* prominently activated *patch* and *white hypopigmentation*. In contrast, body part activations were more sparse and concentrated, with only a few regions such as *head cheeks*, *head scalp*, and *lower extremities thighs* contributing meaningfully. The under-utilization of many anatomical concepts possibly suggests the model’s insensitivity to spatial context.

This selective concept usage underscores the need for additional constraints that promote semantic coverage and balanced concept learning. Future work could incorporate concept entropy regularization or supervision-aware attention mechanisms to encourage more uniform engagement across the concept space, particularly for underrepresented anatomical regions.

C.2 Concept Contribution Analysis Across Semantically Disjoint Families

To further explore the interplay between *descriptor* and *anatomical* concept families within our Concept Bottleneck Model (CBM), we provide two sets of illustrative examples (Sup. Fig. 4, 5). These were chosen from correctly predicted samples with concept annotations verified by expert dermatologists. Contribution scores are computed using signed, log-scaled intermediate logits, reflecting each concept’s influence (positive or negative) on the final class prediction.

(a) Positive Contribution from Body Part Concepts with Co-activation of Descriptors. In Sup. Fig. 4, we show eight representative cases where at least one body part concept has a positive contribution to the model’s decision. In seven of these, we observe concurrent positive contributions from descriptor concepts, as well, suggesting that when anatomical information is utilized by the model, it is rarely used in isolation. Only one sample showed no positively contributing descriptor, which may hint at either spurious localization or weak feature learning from descriptors in that particular case. This consistent pattern of joint activation supports the view that meaningful spatial reasoning in the model emerges most reliably when supported by surface-level lesion features, reinforcing the clinical validity of concept co-activation as a desirable property in interpretable models.

(b) Dominance of Descriptor Concepts in Prediction with Absent Positive Anatomical Contributions. Sup. Fig. 5 complements the above by presenting eight cases where descriptor concepts show strong positive contributions, but none of the body part concepts make a positive impact. While both positively and negatively contributing concepts are visible, the absence of anatomical contributions even in correctly classified examples highlights the model’s stronger reliance on surface-level lesion descriptors.

These observations are consistent with the concept frequency analysis reported earlier, where descriptor concepts not only activated more frequently but also aligned more closely with disease-specific

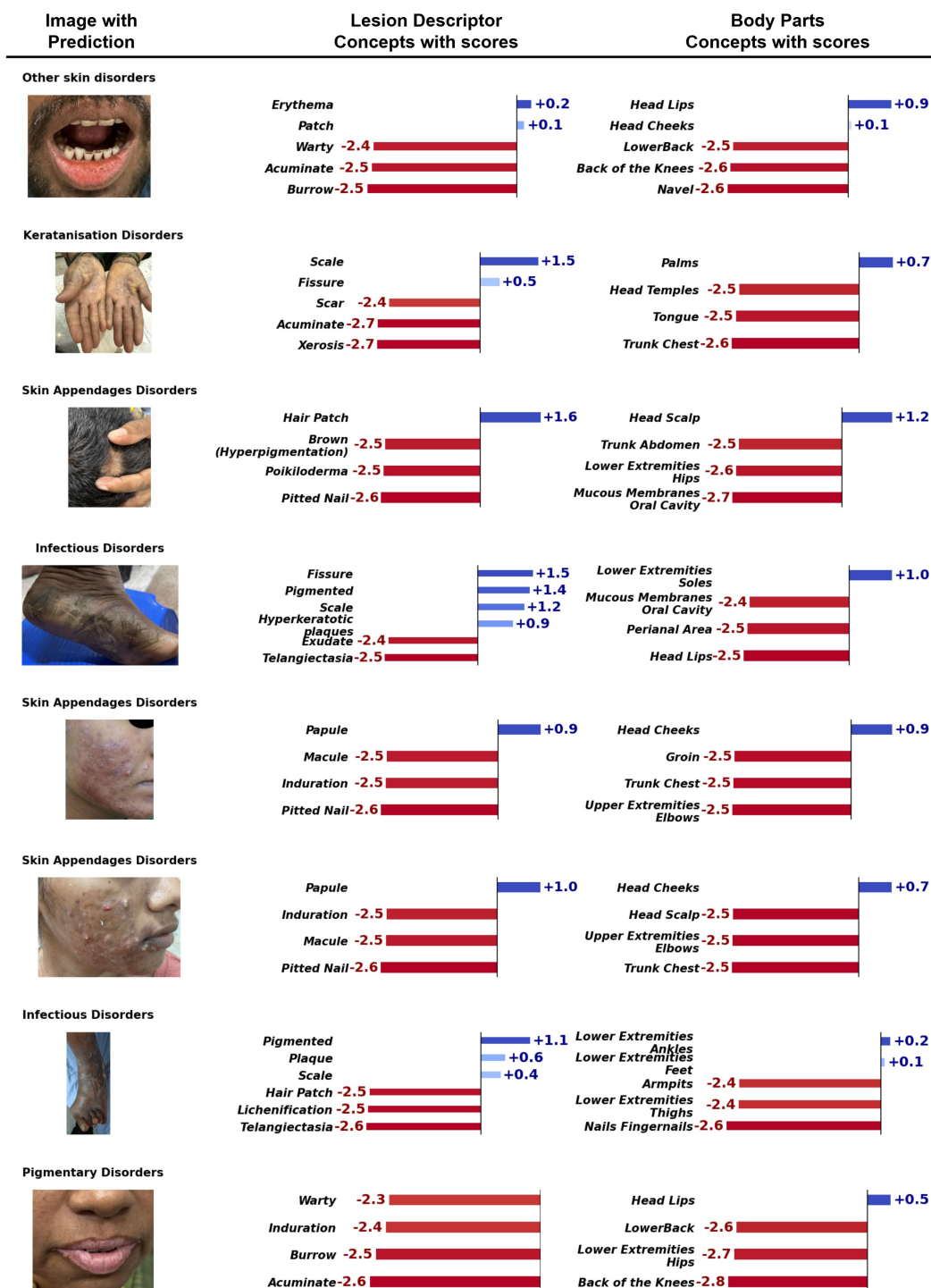


Figure 4: Bar plots showing the top and bottom-k contributing concepts (lesion descriptors and body parts) for the model's prediction. Contribution values are computed as signed, log-scaled scores derived from the CBM's intermediate concept logits. Blue bars indicate the concepts with positive contributions, whereas the red bars highlight the concepts with negative contributions, and all are rightly predicted. These examples specifically highlight cases where at least one body part concept has a positive contribution, with lesion descriptors also showing concurrent scores in most instances.

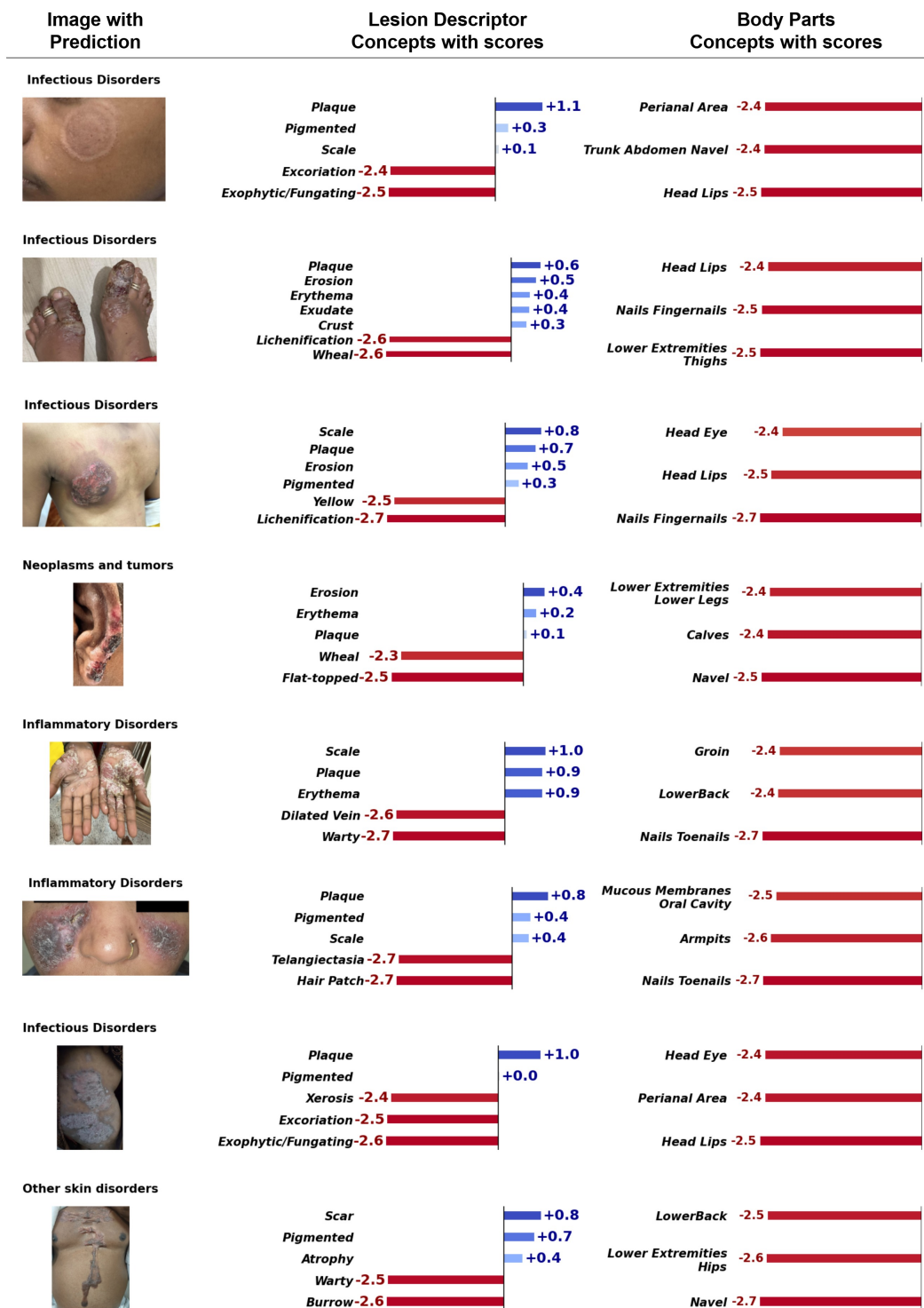


Figure 5: Bar plots showing the top and bottom-k contributing concepts (lesion descriptors and body parts) for the model's prediction. Contribution values are computed as signed, log-scaled scores derived from the CBM's intermediate concept logits. Blue bars indicate the concepts with positive contributions, whereas the red bars highlight the concepts with negative contributions, and all are rightly predicted. These examples specifically highlight cases where at least one lesion descriptor concept has a positive contribution, while body part concepts exhibit negative contribution.

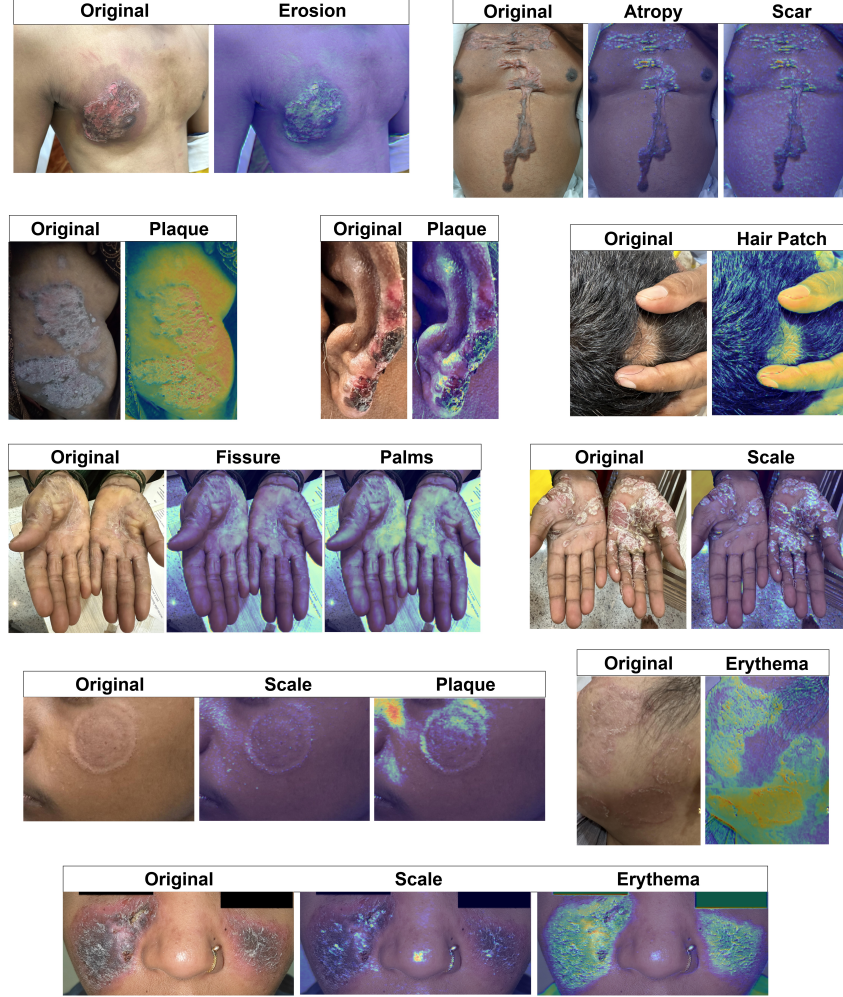


Figure 6: Illustrations of Grad-cam visualizations over Swin Transformer by choosing a specific concept. Highlighted parts of the images with either green, yellow, or red points to the regions responsible for the positive contribution of the concept. All the images were further cross-validated by the dermatologists for the right concept activation.

patterns. Taken together, these findings reveal a clear representational bias in the CBM toward the descriptor concept family, underscoring the need for regularization strategies or balanced supervision to ensure equitable utilization of spatial and morphological information.

C.3 Concept-specific Spatial Attribution via Grad-CAM

To gain spatial interpretability over concept activations, we performed Grad-CAM analysis on the Swin Transformer’s CBM layer for a subset of correctly classified examples. As illustrated in Sup. Fig. 6, each visualization corresponds to a specific concept with positive contribution selected from the CBM’s intermediate bottleneck (e.g., *plaque*, *scale*, *palms*), with ten examples chosen to demonstrate the range and specificity of concept-localized attention. Using logit-directed gradient backpropagation from the concept prediction head, we generated class-discriminative heatmaps that reveal the spatial regions influencing each concept’s activation. These visualizations reinforce the semantic alignment between model representations and clinical reasoning: lesion descriptors such as *erythema* or *plaque* consistently activate in regions of visible inflammation or raised morphology. Importantly, the ability to isolate spatial attributions per concept allows clinicians to verify not just *what* the model has learned, but also *where* it is looking, serving as a crucial step toward validating model trustworthiness in clinical settings.

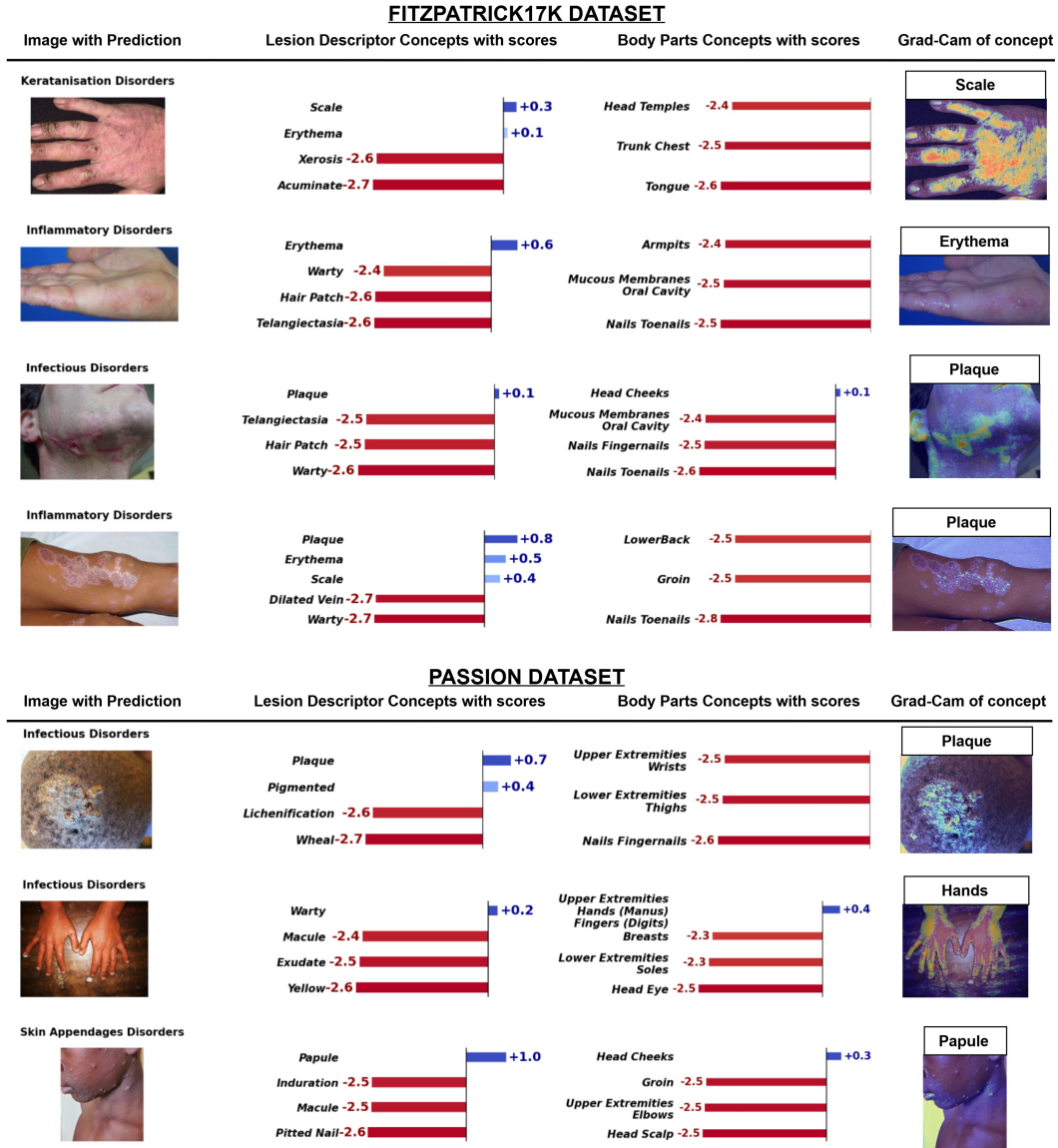


Figure 7: Illustration of interpretability analysis of CBM (Type 2) model trained on DermaCon-IN dataset and cross validated on samples of Fitzpatrick17k and PASSION datasets. Bar plots show the top and bottom-k contributing concepts (lesion descriptors and body parts) for the model's prediction. Contribution values are computed as signed, log-scaled scores derived from the CBM's intermediate concept logits. Blue bars indicate the concepts with positive contributions, whereas the red bars highlight the concepts with negative contributions, and all are rightly predicted. The rightmost column presents the Grad-Cam analysis of the best concept.

D Cross-Dataset Validation and Distributional Coverage

Our dataset has been developed with a focus on flexibility and interoperability, aiming to support a range of downstream dermatological AI tasks. Its hierarchical taxonomy enables researchers to adjust label granularity according to their specific objectives, whether for disease classification, concept prediction, or region-aware modeling.

To demonstrate this integrative potential, we conducted cross-dataset evaluations using two publicly available benchmarks: the PASSION [8] dataset and the Fitzpatrick17k [10] dataset. Using our label hierarchy, we aligned 40 randomly selected samples (due to unavailability of val splits) from each dataset to our taxonomy, as the label space for both datasets was different. The best performing CBM (Type 2) model, trained exclusively on DermaCon-IN dataset, correctly predicted 30 samples from PASSION and 33 from Fitzpatrick17k, and produced clinically valid and interpretable Grad-CAM visualizations across these external samples. All outputs were reviewed and verified by a board-certified dermatologist for both diagnostic accuracy and localisation relevance. A few samples of our analysis is presented in Sup. Fig. 7.

These observations suggest that our dataset is neither isolated nor out-of-distribution. Rather, it addresses a critical representational gap by contributing cases from underrepresented skin tones, outpatient clinical settings, and real-world diagnostic variation specific to South Asian populations.

While further large-scale benchmarking is necessary, our results suggest that the structure and diversity of the dataset provide practical value for researchers dealing with heterogeneous dermatological data. Although we do not claim generalisability, the dataset can be an asset for training models aimed towards transferability and foundational learning.

E Ethical Considerations

This dataset was curated through clinical data collection from consenting patients from outpatient clinics. Ethical diligence was integrated into the dataset lifecycle, including data collection, annotation, privacy protection, and intended use, in line with emerging best practices for responsible dataset curation in human-centric computer vision (HCCV). The protocol was reviewed and approved by the institutional ethics committee.

Data Source and Informed Consent All images were collected during routine dermatological consultations with informed consent from patients in the native language or a language more easily understood by the patients. No data was scraped from the web or obtained from public platforms.

Privacy Protection and Anonymization We ensured that no personally identifiable information (PII) was present in any image and is irreversibly coded for identity. Facial identifiable features such as eyes, tattoos, and other identifiable marks were excluded or cropped from the images. Additionally, all embedded metadata (e.g., timestamps, device IDs, location data) was removed. Participants were informed of their right to withdraw from the study.

Cultural Sensitivity. During image collection involving female patients, all study procedures were explained by, or in the presence of, female medical staff to ensure comfort, privacy, and culturally appropriate engagement.

Annotation Integrity and Labeling Ethics All annotations were carried out by trained clinical annotators working under a dermatologist’s supervision. We deliberately avoided inferring any identity-related attributes from images. Skin tone annotations were made directly based on non-affected skin regions using standardized MST/Fitzpatrick reference scales. Inter-annotator consistency was maintained through calibration sessions, and difficult cases were adjudicated by a senior dermatologist.

Use Restrictions and Responsible Deployment The dataset is intended for academic research, particularly in studying fairness, robustness, and interpretability in dermatological disease classification. It is not intended for commercial use, biometric profiling, identity inference, or deployment in real-time diagnostic tools without regulatory and clinical validation. We intend to make the dataset

openly accessible to support research in dermatological AI. However, access will be granted following a background verification process to ensure responsible and ethical use.

F Societal Impact

The potential impact of this dataset lies not in the scale of data amassed, but in its grounding within everyday clinical reality. By capturing routine skin conditions in South Asian outpatient settings, where diagnostic decisions often occur with limited advanced imaging, the dataset reflects how dermatological care is actually practiced in many parts of the world. Its value, therefore, is not just in building algorithms, but in offering a testbed where machine learning systems can be evaluated under conditions that mirror frontline clinical environments. Additionally, by organizing information through interpretable clinical descriptors and anatomical context, the dataset invites collaboration between clinicians and model developers, not as a retrospective audit, but as an integral part of design. This is particularly meaningful for low-resource settings where human expertise must be complemented. Rather than aiming for universal solutions, we hope this dataset contributes to a shift toward more situated, dialogic approaches to clinical AI development.

Beyond its immediate utility for algorithm training, the dataset also raises possibilities for advancing how uncertainty and variability are treated in clinical AI. By including lesions with atypical or overlapping features, cases that might otherwise be excluded, it encourages models to engage with diagnostic edge cases rather than avoid them. This is particularly relevant in primary care and peripheral settings, where clear-cut presentations are the exception, not the norm. Additionally, the dataset provides opportunities to study how different combinations of features influence clinical suspicion, enabling future work on model calibration and risk stratification. In this way, the dataset serves not only as input to predictive systems and as a scaffold for developing tools that assist with triage, escalation decisions, or patient education — areas where uncertainty is not a flaw but a central part of the task.

G Glossary of Terms

Glossary continued across pages.

Term	Definition
Alterations	Alterations denote clinically observable changes in lesion attributes such as color, texture, or size over time, often indicating progression, regression, or therapeutic response.
Anatomical	Anatomical refers to the specific bodily region where a lesion appears, which can influence diagnostic reasoning due to region-specific disease prevalence and morphology.
Anomalies	Anomalies represent structural or morphological deviations from typical presentation, often indicating pathological states like vascular malformations or congenital disabilities.
Appendageal	Appendageal pertains to skin-associated structures such as hair follicles, sebaceous glands, and nails, diagnostically relevant in conditions like alopecia or onychomycosis.
Atopic	Atopic describes conditions driven by hypersensitivity or allergic predisposition, notably atopic dermatitis, often linked to genetic and environmental triggers.
Blistering	Blistering refers to fluid-filled skin elevations (vesicles/bullae).
Chronic	Chronic refers to a Long-standing or recurrent condition; affects annotation and model expectations.
Corticosteroid	Corticosteroid refers to the Anti-inflammatory drug class; often implicated in misuse or treatment.
Curricula	Curricula refer to Formal educational content; they provide background for how clinicians interpret and label data.
Cutaneous	Relating to the skin; distinguishes from mucosal involvement.

Continued on next page

Continued from previous page

Term	Definition
Demographic	Relating to population traits; crucial for fairness and representation.
Dermatological	About the skin and its diseases; defines the scope of the dataset.
Dermatophytosis	Dermatophytosis refers to a Fungal infection of skin, hair, or nails.
Dermatoses	Dermatoses refers to a general term for skin diseases.
Dermis	Dermis refers to the middle layer of skin, the site of many dermatological processes.
Dermoscopy	Dermoscopy is a non-invasive skin imaging technique for assessing pigmented lesions.
Diagnostic	Related to identifying diseases; informs how images are labeled.
Endemic	Regularly found among a population; informs dataset collection region.
Epidemiologically	In terms of disease patterns in populations; informs dataset design.
Epidermis	Epidermis refers to the outer skin layer, involved in superficial lesions and visual markers.
Erosion	Erosion refers to the loss of part of the epidermis; visible and diagnosable via image.
Etiology	The cause or origin of a disease; critical for understanding disease mechanisms in datasets.
Exogenous	Exogenous refers to originating outside the body; affects classification of environmental skin damage.
Fluctuant	Fluctuant refers to Soft and compressible; indicates the presence of fluid (e.g., cysts, abscesses).
Granulomatous	Related to granuloma formation; chronic inflammation marker.
Hematological	Related to blood or blood-forming organs; systemic diseases may manifest cutaneously.
ICD	ICD refers to the International Classification of Diseases; a standardized coding system used globally for diagnosis and reporting.
Incontinence	In dermatology, refers to pigment incontinence where melanin leaks into the dermis.
Infectious	Caused by pathogens; represents a major disease category in dermatology.
Inflammatory	Involving immune response; another common skin disease category.
Keratinization	Keratinization refers to skin thickening; central in conditions like psoriasis.
LMIC	LMIC refers to Low- and Middle-Income Countries; refers to geographic and economic contexts.
Lesion	Lesion refers to any abnormal area on the skin, the primary subject of dermatological datasets.
Melanin	Melanin refers to a pigment responsible for skin color; key in diagnosing pigmentation disorders.
Melanoma	Melanoma refers to a malignant tumor of melanocytes, a key example of a neoplastic skin lesion.
Morbidity	Morbidity refers to the rate of disease in a population; it helps contextualize dataset relevance.
Morphological	Morphological refers to Concerned with form and structure; describes visual features of lesions.
Mucosal	Mucosal refers to moist linings (e.g., inside mouth); important in systemic diseases.
Neoplastic	Neoplastic refers to abnormal cell growth; includes benign and malignant tumors.
Ontologies	Ontologies refer to Structured vocabularies linking concepts; useful for clinical data labeling.
Outpatient	Outpatient refers to a Healthcare setting where patients are not admitted overnight; a common source of dermatology images.
Palpable	Able to be felt; clinical term for raised or solid lesions.
Pathological	Related to disease processes; central to diagnosis and data labeling.
Pathophysiology	Pathophysiology refers to functional changes associated with disease; it links imaging to mechanisms.

Continued on next page

Continued from previous page

Term	Definition
Phenotypic	Phenotypic refers to observable traits or characteristics; crucial for image-based annotations.
Pigment	Pigment refers to coloring matter in skin; changes indicate various disorders such as vitiligo or melasma.
Pigmentary	Pigmentary refers to relating to skin color changes; includes hyper- and hypo-pigmentation.
Pigmentary Alterations	Pigmentary Alterations refers to Changes in skin color, important for diagnosing pigmentation disorders.
Polymorphic	Polymorphic refers to having varied forms; describes diverse visual patterns of lesions.
Primary Lesions	Primary Lesions refers to Initial skin changes (e.g., macule, papule); basis for clinical diagnosis.
Pruritic	Pruritic refers to Itchy; common symptom that guides diagnosis.
Rook's	Rook's refers to Rook's Textbook of Dermatology; a key reference in dermatological classification and clinical teaching.
Secondary Changes	Secondary Changes refer to lesion alterations due to disease progression or external factors.
Subcutaneous	Subcutaneous refers to the layer beneath the dermis composed of fat and connective tissue; affected in deep infections or nodules.
Systemic	Systemic refers to affecting the entire body; differentiates skin manifestations of internal diseases.
Taxonomies	Taxonomies refer to systematic classification of concepts or conditions; essential for dataset organization and ontology mapping.
Varicosities	Varicosities refer to dilated veins; visible skin features relevant in elderly or vascular conditions.
Vascular	Vascular refers to relating to blood vessels; includes lesions like purpura and telangiectasia.
Vasculitis	Vasculitis refers to the Inflammation of blood vessels; it can present with palpable purpura.
Vesiculobullous	Vesiculobullous refers to Diseases characterized by vesicles and bullae (e.g., pemphigus).
Xanthomas	Xanthomas refers to Lipid-rich lesions; indicative of metabolic disorders.
Xerotic	Xerotic refers to Dry skin; common in aging and environmental dermatitis.

References

- [1] N. Alipour et al. Skin type diversity in skin lesion datasets: A review. *International Journal of Dermatology*, 63:198–210, 2024. doi:10.1007/s13671-024-00440-0.
- [2] P. Balasubramanian et al. Epidemiological study of skin disorders in andaman and nicobar islands. *Indian Journal of Dermatology*, 66(5):454–458, 2021. doi:10.4103/ijid.IJD_30_20.
- [3] Jean L. Bolognia, Julie V. Schaffer, and Lorenzo Cerroni. *Dermatology*. Elsevier, 4 edition, 2017. ISBN 9780702062759.
- [4] R. Daneshjou et al. Disparities in dermatology ai performance on a diverse, curated clinical image set. *Scientific Reports*, 12:12565, 2022. doi:10.1126/sciadv.abq6147.
- [5] R. Daneshjou et al. Skincon: A skin disease dataset densely annotated by domain experts for fine-grained debugging and analysis. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/7318b51b52078e3af28197e725f5068a-Abstract-Datasets_and_Benchmarks.html.
- [6] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.
- [7] Thomas B. Fitzpatrick. Soleil et peau. *Journal de Médecine Esthétique*, 2:33–34, 1975.
- [8] P. Gottfrois et al. Passion for dermatology: Bridging the diversity gap with pigmented skin images from sub-saharan africa. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pages 703–712. Springer Nature Switzerland, 2024.
- [9] Christopher E.M. Griffiths et al. *Rook’s Textbook of Dermatology*. Wiley-Blackwell, 10 edition, 2024. ISBN 9781119709213. URL <https://www.wiley.com/en-us/Rook%27s%2BTextbook%2Bof%2BDermatology%2C%2B4%2BVolume%2BSet%2C%2B10th%2BEdition-p-00402062>. 4-volume set.
- [10] M. Groh et al. Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1820–1828, 2021.
- [11] Robert J Hay, Neil E Johns, Hannah C Williams, and et al. The global burden of skin disease in 2010: An analysis from the global burden of disease study 2010. *The Journal of Investigative Dermatology*, 134(6):1527–1534, 2014. doi:10.1038/jid.2013.446.
- [12] Roderick J. Hay et al. Skin disease in the tropics and the lessons that can be learned from leprosy and other neglected diseases. *Acta Dermato-Venereologica*, 100:adv00113, 2020. doi:10.2340/00015555-3469.
- [13] Kaiming He et al. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [14] C. Hernández et al. Bcn20000: Dermoscopic lesions in the wild. *Scientific Data*, 11:641, 2024. doi:10.1038/s41597-024-03387-w.
- [15] Pengcheng Huai et al. Global burden of skin and subcutaneous diseases: an update from the global burden of disease study 2021. *British Journal of Dermatology*, lja071, 2025. doi:10.1093/bjd/lja071. Published: 11 April 2025.
- [16] Gao Huang et al. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [17] ISIC. Siim-isic 2020 challenge dataset, 2020. URL <https://doi.org/10.34970/2020-ds01>. Creative Commons Attribution-Non Commercial 4.0 International License.
- [18] N. S. Jayanthi et al. Epidemiological pattern of skin diseases among patients attending dermatological outpatient department at a tertiary care centre, north chennai. *Indian Journal of Clinical and Experimental Dermatology*, 3(4):134–137, 2017. doi:10.18231/2455-6769.2017.0032.

- [19] Emily Johnson and Robert Lee. Global burden of skin diseases in 2019: Updated estimates from the global burden of disease study 2019. *The Lancet Global Health*, 8(11):e1539–e1540, 2020. doi:10.1016/S2214-109X(20)30386-7.
- [20] C. Karimkhani et al. Global skin disease morbidity and mortality: An update from the global burden of disease study 2013. *JAMA Dermatology*, 153(5):406–412, 2017. doi:10.1001/jamadermatol.2016.5538.
- [21] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh. 7-point checklist and skin lesion classification using multi-task multi-modal neural nets. *IEEE Journal of Biomedical and Health Informatics*, Apr 2018. doi:10.1109/JBHI.2018.2824327. Epub ahead of print.
- [22] Pang Wei Koh et al. Concept bottleneck models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 5338–5348, 2020.
- [23] Melissa R. Laughter, Mayra B. C. Maymone, Chante Karimkhani, Chandler Rundle, Sophia Hu, Sophia Wolfe, Katrina Abuabara, Parker Hollingsworth, Gil S. Weintraub, Cory A. Dunnick, Adnan Kisa, Giovanni Damiani, Aziz Sheikh, Jasvinder A. Singh, Takeshi Fukumoto, Rupak Desai, Ayman Grada, Irina Filip, Amir Radfar, Mohsen Naghavi, and Robert P. Dellavalle. The burden of skin and subcutaneous diseases in the united states from 1990 to 2017. *JAMA Dermatology*, 156(8):874–881, 2020. doi:10.1001/jamadermatol.2020.1573. URL <https://jamanetwork.com/journals/jamadermatology/fullarticle/2767074>.
- [24] Ze Liu et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.
- [25] M. López-Pérez et al. Are generative models fair? a study of racial bias in dermatological image generation, 2025. URL <https://arxiv.org/abs/2501.11752>.
- [26] T. Mendonça et al. Ph²: A dermoscopic image database for research and benchmarking. In *Proceedings of the 35th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, July 2013.
- [27] A. O’Malley et al. Ensuring appropriate representation in artificial intelligence-generated medical imagery: Protocol for a methodological approach to address skin tone bias. *JMIR AI*, 3:e58275, 2024. doi:10.2196/58275. URL <https://ai.jmir.org/2024/1/e58275>.
- [28] A. G. C. Pacheco et al. Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, 32:106221, 2020. doi:10.1016/j.dib.2020.106221.
- [29] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. <https://github.com/pytorch/pytorch>, 2019.
- [30] Ronald P. Rapini. *Dermatology: 2-Volume Set*. Mosby, 1 edition, 2007. ISBN 9780721601573.
- [31] Candice Schumann, Femi Olanubi, Auriel Wright, Ellis Monk, Courtney Heldreth, and Susanna Ricco. Consensus and subjectivity of skin tone annotation for ml fairness. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023), Datasets and Benchmarks Track*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/60d25b3210c92f5ba2002a8e1f1adf1c-Abstract-Datasets_and_Benchmarks.html.
- [32] B. Shah et al. Epidemiological study of skin diseases in himatnagar. *International Journal of Research in Dermatology*, 5(2):342–345, 2019. doi:10.18203/issn.2455-4529.IntJResDermatol20190453.
- [33] X. Sun et al. A benchmark for automatic visual classification of clinical skin disease images. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, volume 9910 of *Lecture Notes in Computer Science*, pages 206–222. Springer, Cham, 2016. doi:10.1007/978-3-319-46466-4_13.

- [34] M. D. Szeto et al. Dermatologic data from the global burden of disease study 2019 and the patientslikeme online support community: Comparative analysis. *JMIR Dermatology*, 7:e50449, 2024. doi:10.2196/50449. URL <https://pubmed.ncbi.nlm.nih.gov/39661989/>.
- [35] Mingxing Tan et al. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 6105–6114. PMLR, 2020.
- [36] K. J. S. Thakur et al. The burden of skin diseases in india: Global burden of disease study 2017. *Indian Journal of Dermatology, Venereology and Leprology*, 87(6):764–771, 2021. doi:10.25259/IJDVL_978_20.
- [37] P. Tschandl et al. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5:180161, 2018. doi:10.1038/sdata.2018.161. URL <https://doi.org/10.1038/sdata.2018.161>.
- [38] Zihang Tu et al. Maxvit: Multi-axis vision transformer. In *European Conference on Computer Vision (ECCV)*, 2022.
- [39] Katelyn Urban et al. The global, regional, and national burden of fungal skin diseases in 195 countries and territories: A cross-sectional analysis from the global burden of disease study 2017. *JAAD International*, 2:22–27, 2021. doi:10.1016/j.jdin.2020.10.003.
- [40] A. Ward et al. Creating an empirical dermatology dataset through crowdsourcing with web search advertisements. *JAMA Network Open*, 7(11):e2446615, November 2024. doi:10.1001/jamanetworkopen.2024.46615.
- [41] World Health Organization. *International Classification of Diseases, 11th Revision (ICD-11)*, 2019. URL <https://icd.who.int/>. Adopted by the 72nd World Health Assembly in 2019; came into effect on 1 January 2022.
- [42] Aobuliximu Yakupu et al. The burden of skin and subcutaneous diseases: findings from the global burden of disease study 2019. *Frontiers in Public Health*, 11:1145513, 2023. doi:10.3389/fpubh.2023.1145513.
- [43] A. Yilmaz et al. Derm12345: A large, multisource dermatoscopic skin lesion dataset with 40 subclasses. *Scientific Data*, 11(1):1302, November 2024. doi:10.1038/s41597-024-04104-3.