nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-020-0842-3

In the format provided by the authors and unedited.

A deep learning system for differential diagnosis of skin diseases

Yuan Liu[1], Ayush Jain[1], Clara Eng[1], David H. Way[1], Kang Lee[1], Peggy Bui[1,2], Kimberly Kanada[3], Guilherme de Oliveira Marinho[4], Jessica Gallegos[1], Sara Gabriele[1], Vishakha Gupta[1], Nalini Singh[1,5], Vivek Natarajan[1], Rainer Hofmann-Wellenhof[6], Greg S. Corrado[1], Lily H. Peng[1], Dale R. Webster[1], Dennis Ai[1], Susan J. Huang[3], Yun Liu[1]✉, R. Carter Dunn[1,7] and David Coz[1,7]

[1]Google Health, Palo Alto, CA, USA. [2]University of California, San Francisco, San Francisco, CA, USA. [3]Advanced Clinical, Deerfield, IL, USA. [4]Adecco Staffing, Santa Clara, CA, USA. [5]Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Medical University of Graz, Graz, Austria. [7]These authors contributed equally: R. Carter Dunn, David Coz. ✉e-mail: liuyun@google.com

Supplementary Information

Supplementary Methods

Labeler onboarding and certification

In addition to formal board certification, all study participants (dermatologists, PCPs, and NPs) underwent an onboarding process to familiarize with the grading tools. In particular, the dermatologists comprising the reference standard graded 147 cases randomly sampled from the development set as an assessment to ensure consistent grading. For each case, their leading diagnosis was compared to the aggregated opinion of a panel of three experienced U.S. board-certified dermatologists, and only dermatologists who had an top-3 accuracy exceeding 60% participated in determining the reference standard for the validation set. This threshold was chosen based on the statistics of dermatologist grader accuracy, so as to leave room for disagreement in complex cases while ensuring a minimum consistency in grading following guidelines (e.g. specificity of diagnoses) and familiarity with the tool. Among 53 dermatologists who completed the test, the average score was 69%. Three dermatologists did not meet the 60% threshold; none of these dermatologists graded the validation set.

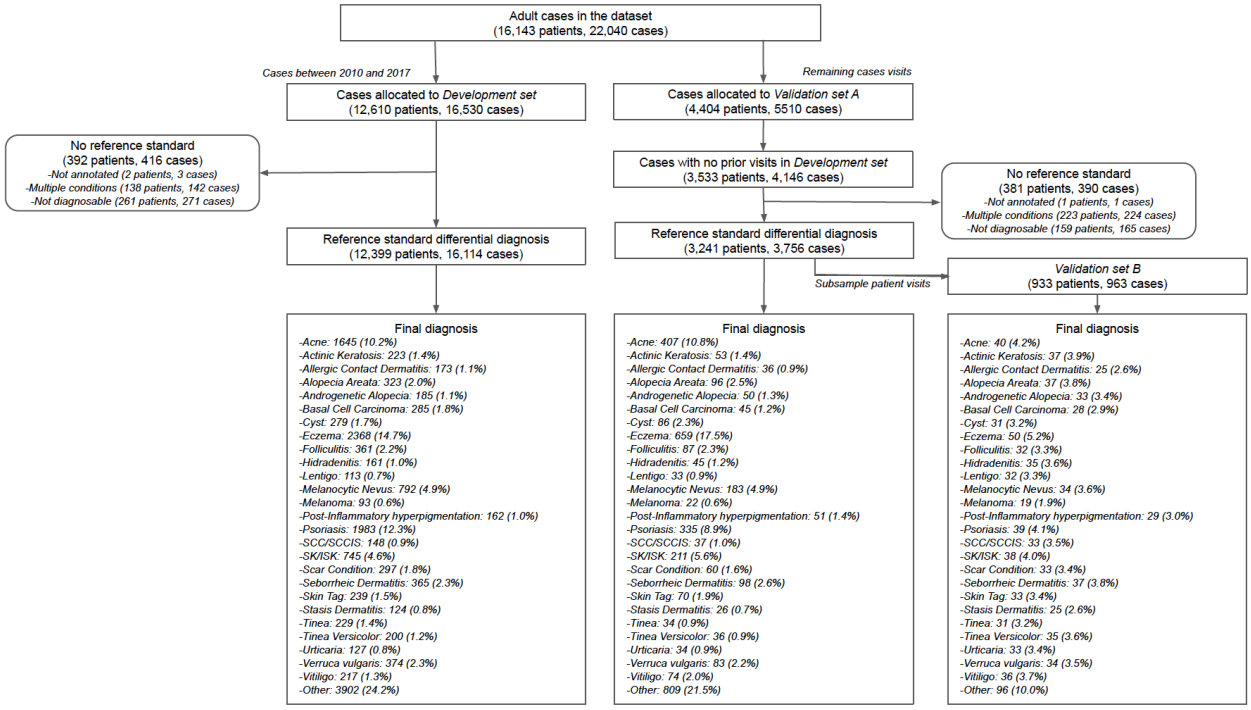
Reference standard voting procedure and reproducibility

Here, we detail the voting procedure[47] used to improve reproducibility of our reference standard (Supplementary Fig. 10). First, each dermatologist provided up to three differential diagnoses and accompanying confidence values in the range [1, 5] for each of the diagnosis. Next, each diagnosis was mapped to a condition. If duplicates occurred (i.e. multiple diagnoses were mapped to the same condition), the highest confidence was retained. The relative ranks of the mapped conditions were used to rank the conditions into a differential diagnosis (i.e. primary, secondary, and tertiary diagnosis). Each mapped condition was then assigned a weight: the inverse of the rank. If multiple mapped conditions shared the same confidence, then the weight was evenly distributed across the conditions. Answers from the dermatologists were then aggregated to form the reference standard, by summing up the weights, before limiting the skin condition classes to 27 (26 conditions plus “Other”) and normalizing their weights to sum to 1. Distribution of the number of conditions in the differential diagnosis for each set is shown in Supplementary Fig. 8. Detailed analysis of the secondary and tertiary diagnoses that are provided alongside every primary diagnosis is shown in Supplementary Fig. 9.

To investigate the reproducibility of the reference standard, for validation set B, three other random dermatologists from the same pool (who had not seen the case

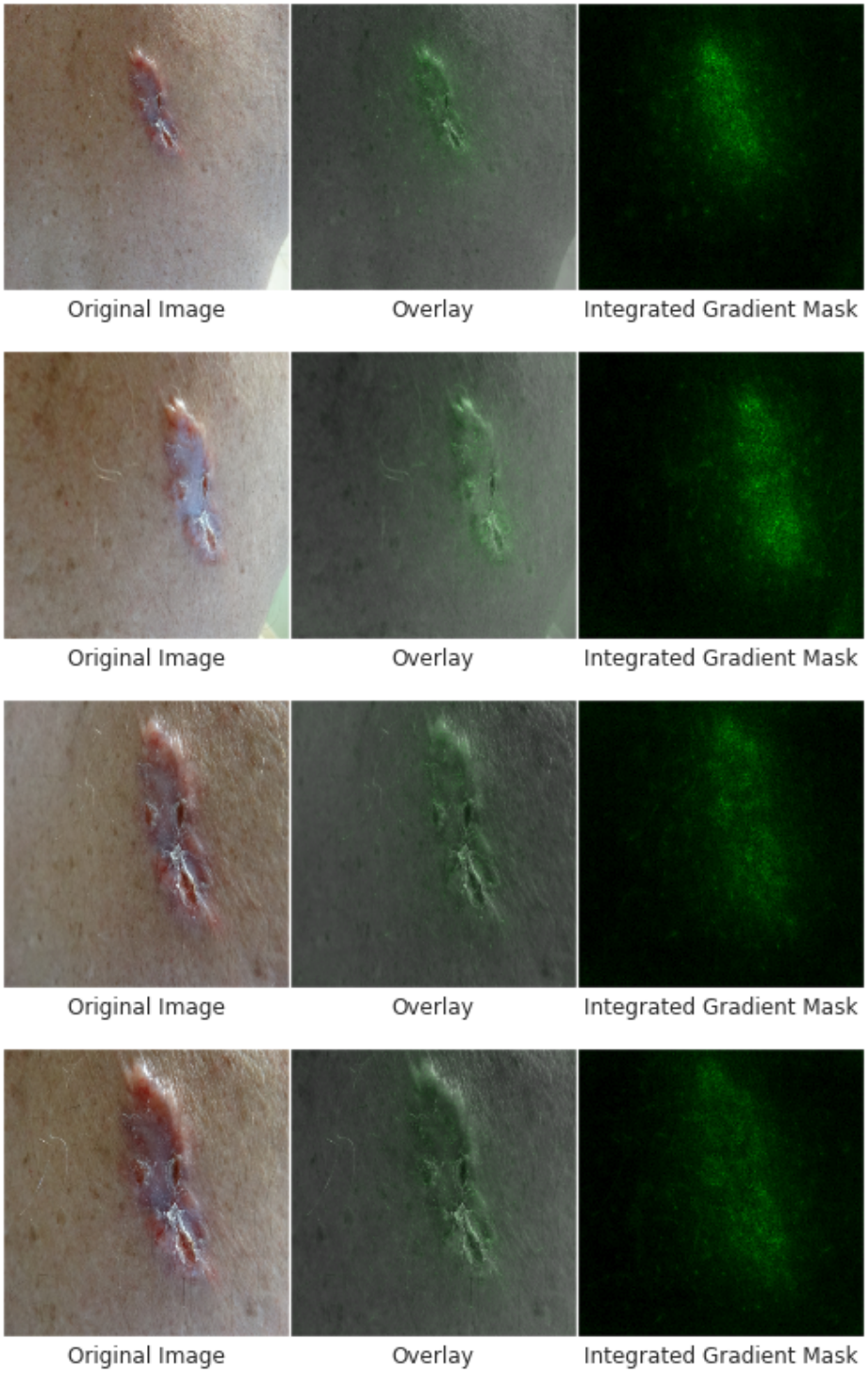
previously) graded the cases independently, following the exact same labeling procedure as before. Reference standard differential diagnoses from the two panels of three dermatologists had an AO of 0.63 and an agreement of 0.74 for the primary diagnosis (compared to an AO of 0.54 and an agreement of 0.61 between two random individual dermatologists, one per panel), when considered in the space of 419 mapped conditions. Within the space of 27 conditions handled by DLS, the two panels had an AO of 0.70 and an agreement of 0.77 (compared to an AO of 0.60 and an agreement of 0.66 between two random individual dermatologists, one per panel).

Supplementary Figures



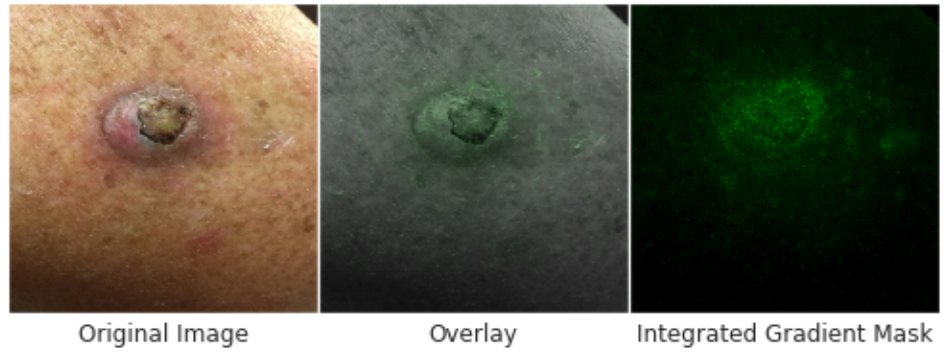
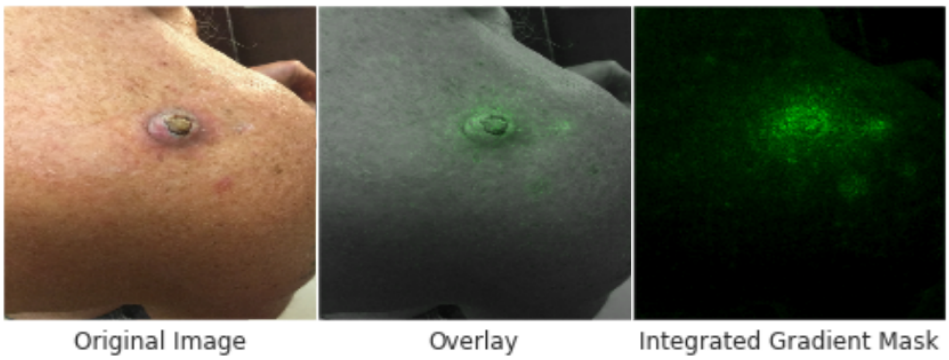
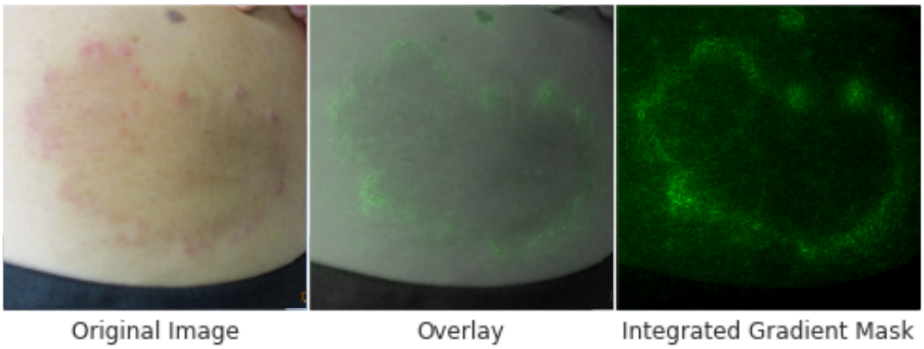
Supplementary Fig. 1 | STARD diagram illustrating the flow of cases used in this work. Patient counts do not add up perfectly because removal of cases only removes patients if no other cases from those patients remain. A small number of cases were not annotated due to technical issues (3 cases in the development set and 1 case in the validation set).

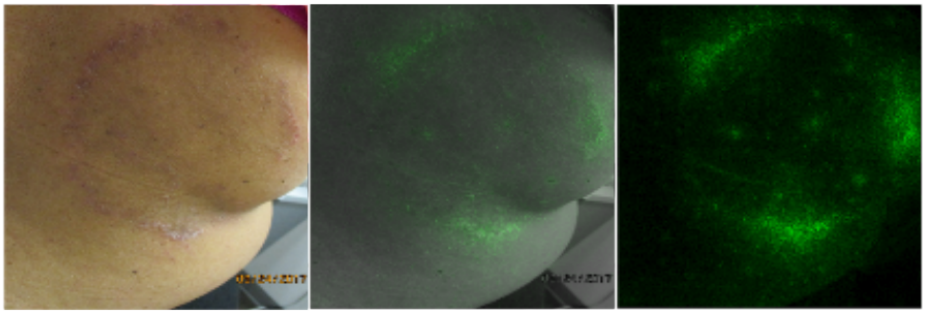
a

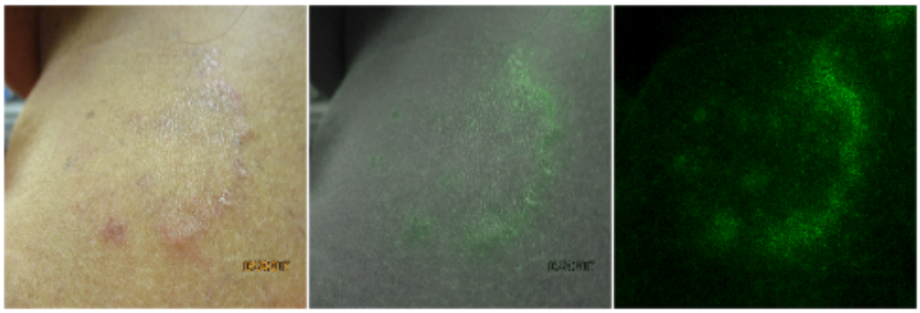


59 year old (y.o.) Male, White
Self-reported skin problem: Rash
Duration: More than one year, always present
Symptoms: Bothersome in appearance, bleeding, increasing in size
Review of system (ROS): No fever/chills (F/C), fatigue, joint pain, mouth sores, or shortness of breath
Drugs: Treated by prescription (Rx) or over-the-counter (OTC)
Medical history: No history of skin cancer, melanoma, eczema, psoriasis, or biopsy
Family history: Psoriasis
Drug allergies: None
Medication: None
Follow-up case?: No

Reference standard	DLS (top 3)	DLS (growth subgroup)	NP (missed)	NP (missed)	PCP (tied 1st diagnosis)	PCP (missed)	Derm	Derm

BCC; SCC/SCCIS / Scar condition	BCC: 0.57; Scar condition: 0.29; SCC/SCCIS: 0.07	Malignant: 0.69; Benign: 0.00	Other (hypertrophic skin); Scar condition	AK; Other (skin lesion); Psoriasis	BCC / SCC/SCCIS; Melanoma	Psoriasis	BCC	BCC
b   						63 y.o. Male Self-reported skin problem: Growth or Mole Duration: Three to twelve months, always present Symptoms: Increasing in size, itching, burning, painful ROS: No F/C, fatigue, joint pain, mouth sores, or shortness of breath Drugs: Treated by Rx or OTC Medical history: No history of skin cancer, melanoma, eczema, psoriasis, or biopsy Family history: Skin cancer Drug allergies: None Medication: None Follow-up case?: No		
Reference standard	**DLS (top 3)**	**DLS (growth subgroup)**	**NP (2nd diagnosis)**	**NP (tied 1st diagnosis)**	**PCP (missed)**	**PCP (missed)**	**Derm**	**Derm**
SCC/SCCIS; BCC	SCC/SCCIS: 0.69; BCC: 0.19; AK: 0.07	Malignant: 0.93; Benign: 0.07	BCC; SCC/SCCIS; Melanoma	Other (skin lesion) / SCC/SCCIS; BCC	Cannot diagnose	Other (pyoderma)	SCC/SCCIS; BCC	SCC/SCCIS
c  						61 y.o. Female, Hispanic or Latino Self-reported skin problem: Rash Duration: One to four weeks, always present Symptoms: Itching ROS: No F/C, fatigue, joint pain, mouth sores, or shortness of breath Drugs: Treated by Rx or OTC Medical history: No history of skin cancer, melanoma, eczema, psoriasis, or biopsy Family history: Skin cancer Drug allergies: None Medication: None		

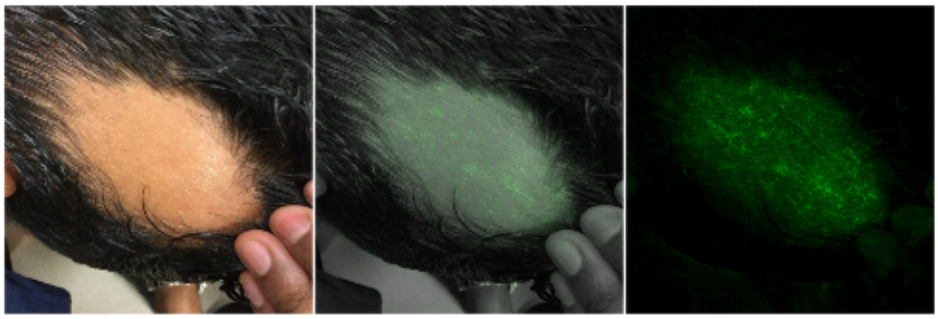




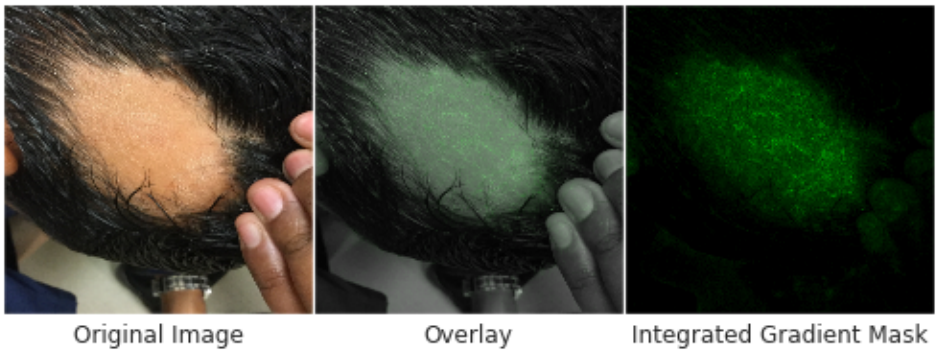
Follow-up case?: No

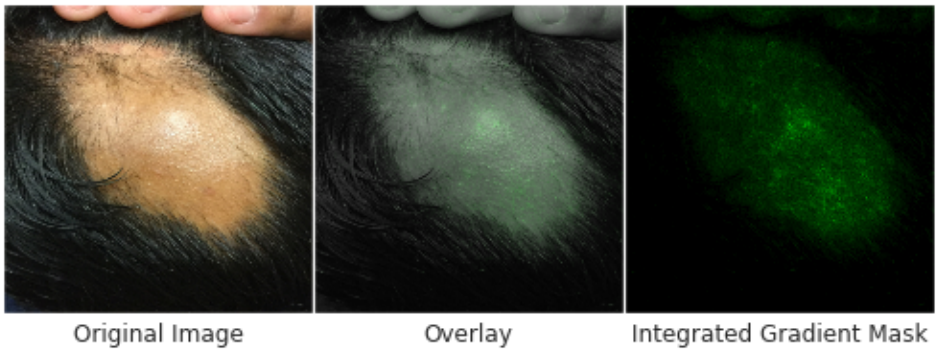
Reference standard	DLS (top 3)	DLS (erythematosquamous and papulosquamous subgroup)	NP (missed)	NP (missed)	PCP (tied 1st diagnosis)	PCP (missed)	Derm	Derm
Tinea	Tinea: 0.68; Eczema: 0.14; Other: 0.03	Infectious: 0.80; Non-infectious: 0.16	Eczema / Other (Chronic contact dermatitis); Psoriasis	Other (Generalized granuloma annulare)	Other (Granuloma annulare) / Tinea	Eczema	Tinea; Other (Granuloma annulare)	Tinea

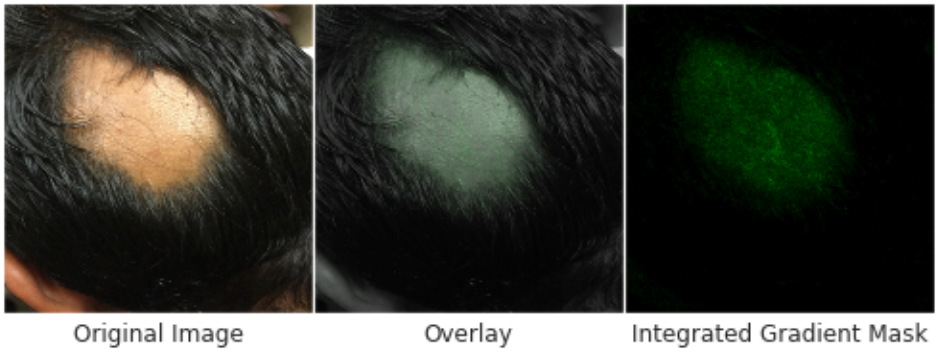
d



29 y.o. Male, Native Hawaiian or Pacific Islander
Duration: Three to twelve months, always present
Symptoms: Bothersome in appearance, increasing in size, itching
ROS: No F/C, fatigue, joint pain, mouth sores, or shortness of breath
Drugs: Treated by Rx or OTC
Medical history: No history of skin cancer, melanoma, eczema, psoriasis, or biopsy
Family history: Skin cancer
Drug allergies: None
Medication: None
Follow-up case?: Yes



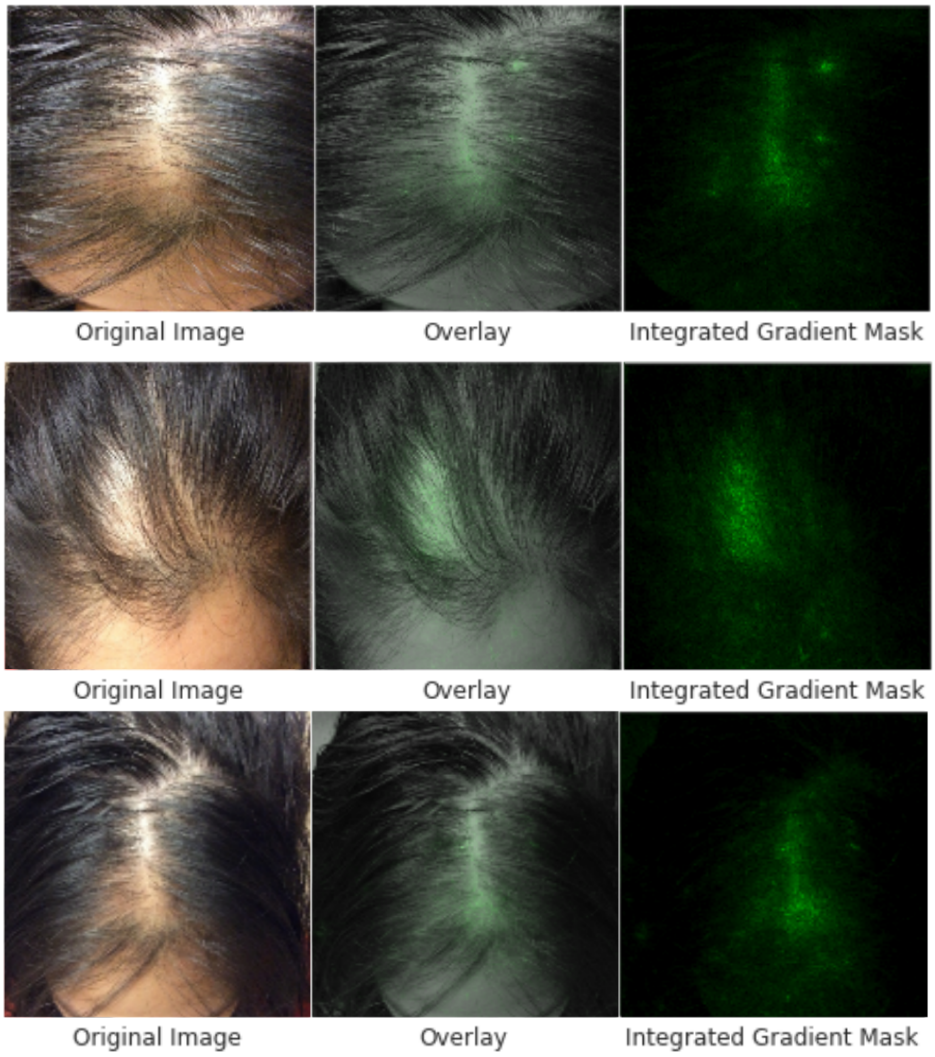




Patient adhered to treatments: Yes
Condition after treatments: Not changed

Reference standard	DLS (top 3)	DLS (hair loss subgroup)	NP (3rd diagnosis)	NP (missed)	PCP (2nd diagnosis)	PCP	Derm	Derm
AA	AA: 0.98; Other: 0.01; AGA: 0.01	AA: 0.98; AGA: 0.01	AGA; Other (Alopecia localis); AA	AGA	AGA; AA	AA	AA	AA; Other (trichotillomania)

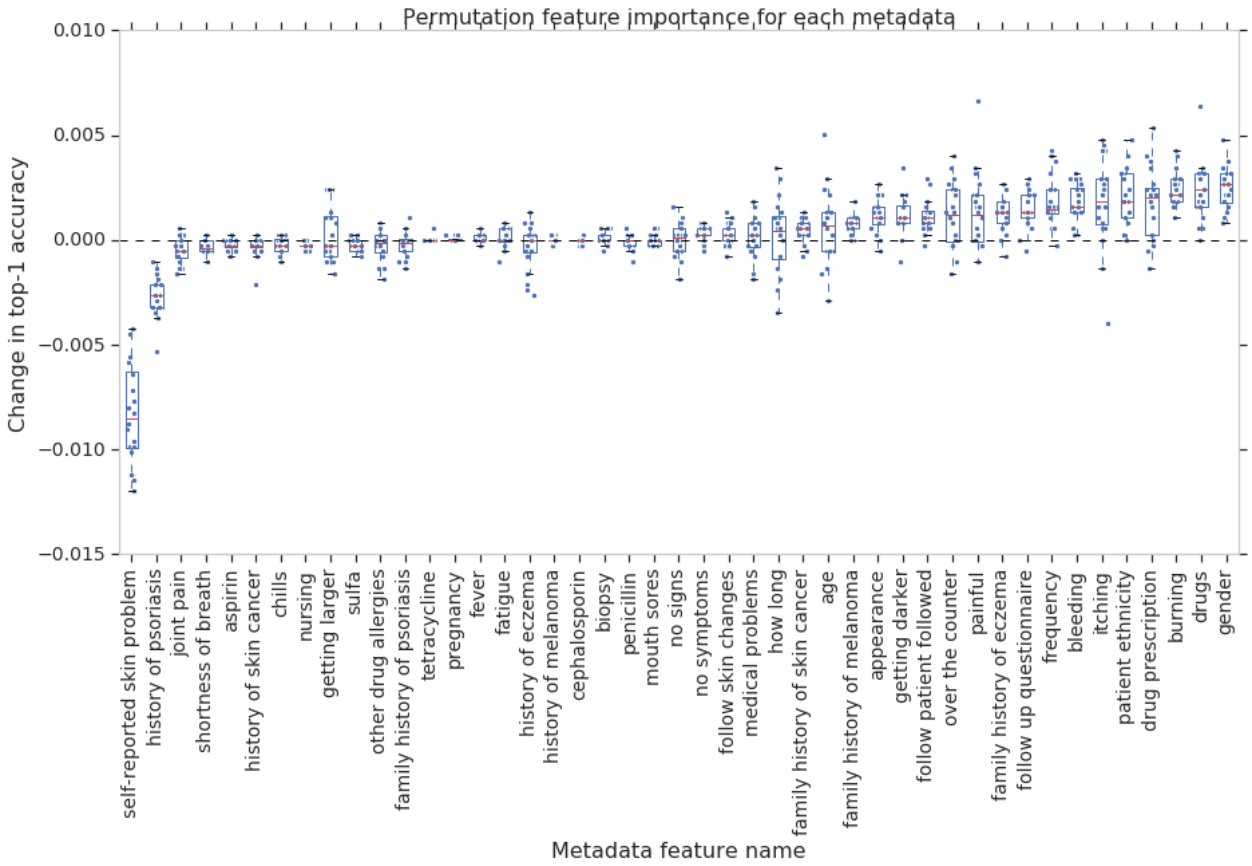
e



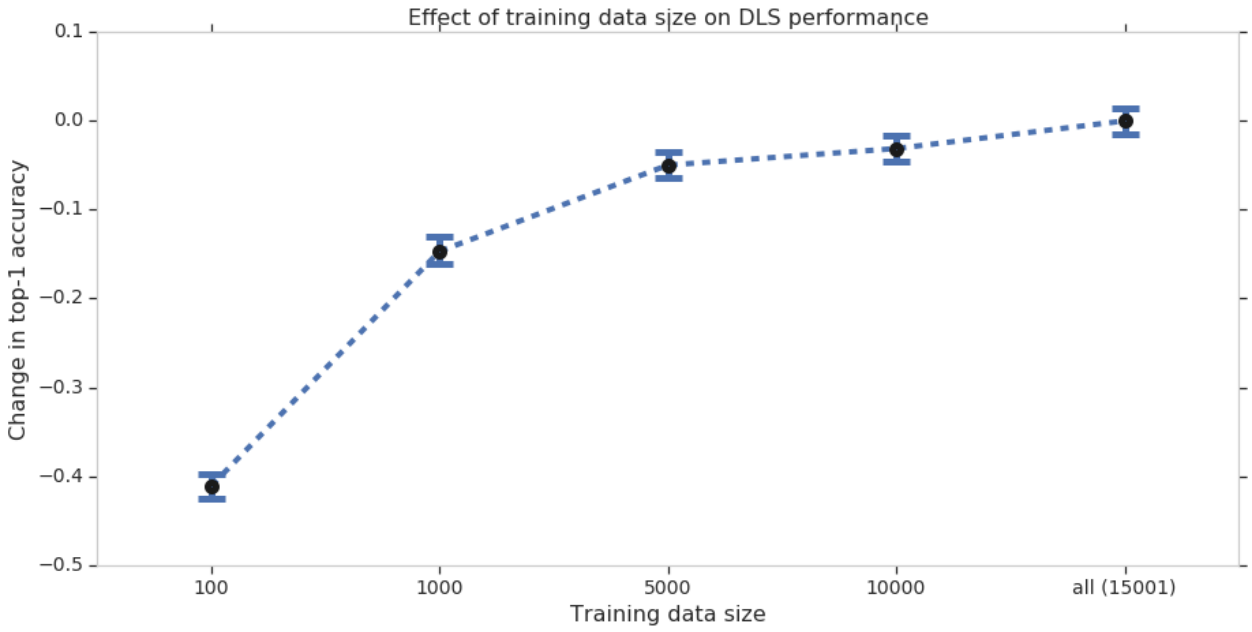
26 y.o. Female, Hispanic or Latino
Self-reported skin problem: Hair loss
Duration: More than five years, always present
Symptoms: Bothersome in appearance, itching
ROS: No F/C, fatigue, joint pain, mouth sores, or shortness of breath
Drugs: Has not been treated by Rx or OTC
Medical history: No history of skin cancer, melanoma, eczema, psoriasis, or biopsy
Family history: Skin cancer, eczema
Drug allergies: None
Medication: OTC
Follow-up case?: No

Reference standard	DLS (top 3)	DLS (hair loss subgroup)	NP (missed, non-specific)	NP (missed, non-specific)	PCP (tied 2[d] diagnosis)	PCP (missed)	Derm	Derm
AGA; AA; Other	AGA: 0.73; Other: 0.13; Seborrheic Dermatitis: 0.08	AGA: 0.78; AA: 0.00	Other (Diffuse alopecia)	Other (Alopecia)	Other (Telogen effluvium); AA / AGA	AA	AGA; Other (Drug-related alopecia)	AGA

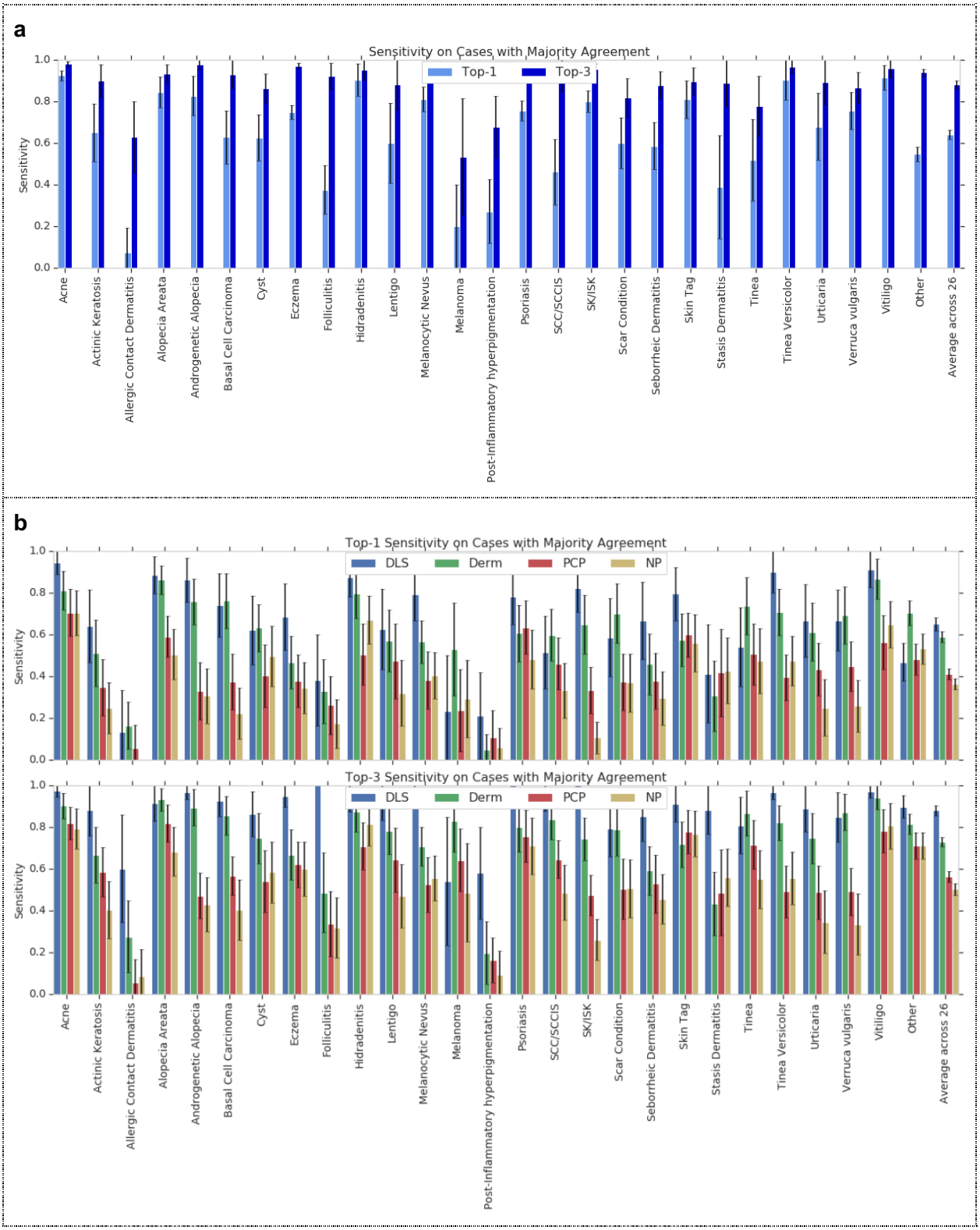
Supplementary Fig. 2 | All the images and metadata for examples shown in Fig. 3. Abbreviations for diagnoses follow those from Fig. 3: basal cell carcinoma (BCC), squamous cell carcinoma (SCC/SCCIS), Alopecia Areata (AA), and Androgenetic Alopecia (AGA). Some images were cropped to zoom in on the condition for clarity.



Supplementary Fig. 3 | Importance of each individual clinical metadata to the deep learning system (DLS). For each clinical metadata, its values are permuted across validation set A examples, and the effect of this permutation on the top-1 accuracy using the same trained DLS are shown. Boxplot meanings are identical to Fig. 4a (which shows only the top 10 features here).

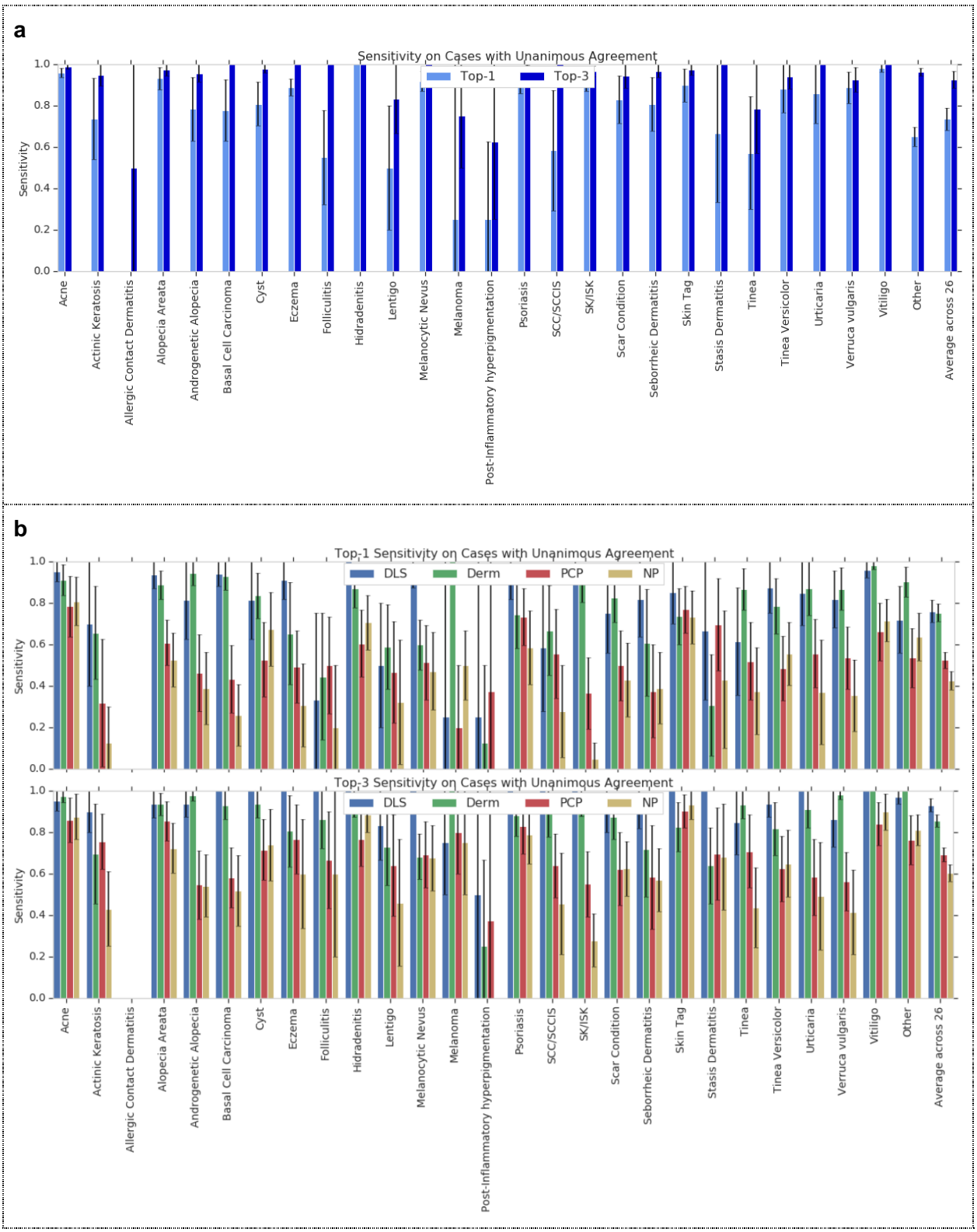


Supplementary Fig. 4 | Effect of training dataset size (excluding tune set) on the performance of the deep learning system (DLS). For each experiment, a random subset of the cases was used for training. This DLS was then evaluated on the validation set A and its change in the top-1 accuracy relative to the original DLS (trained with all available training data) is shown. Error bars indicate 95% confidence intervals across all cases in validation set A (n=3,756).



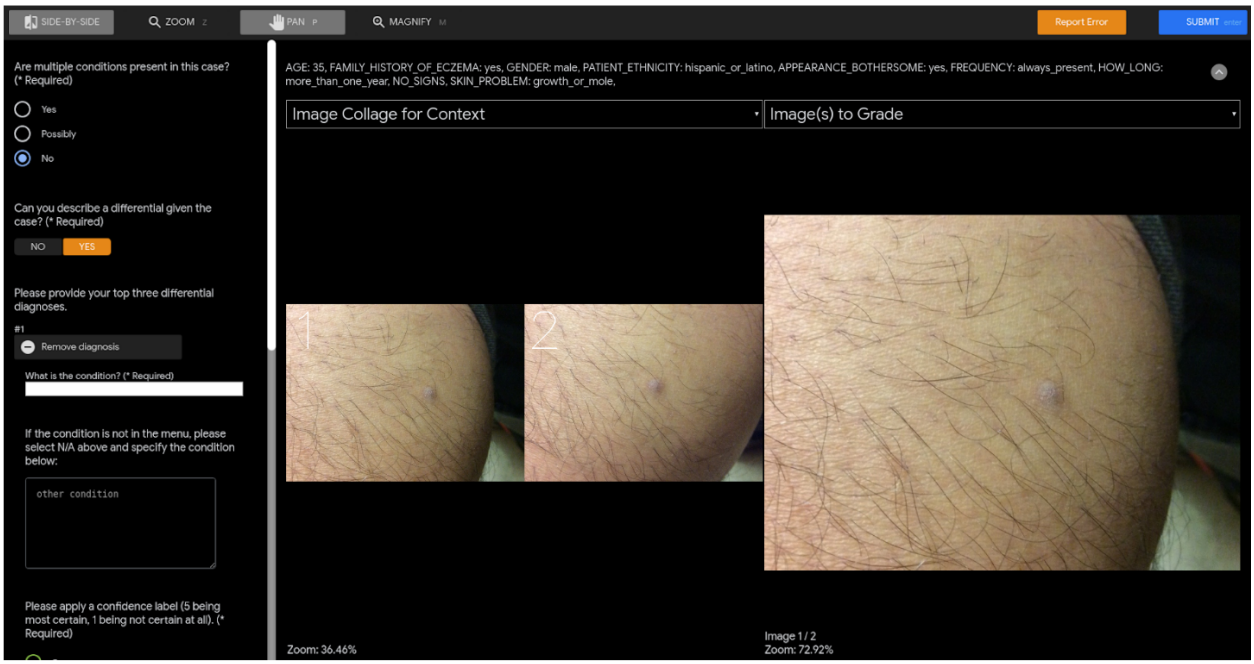
Supplementary Fig. 5 | Performance of the deep learning system (DLS) and clinicians in cases where at least two out of the three dermatologists determining the reference standard agreed on the primary diagnosis, broken down for each of the 26 categories of skin conditions. a, Top-1 and top-3 sensitivity of the DLS on validation set A (n=3,756). **b**,

Top-1 and top-3 sensitivity of the DLS and three types of clinicians: dermatologists (Derm), primary care physicians (PCP), and nurse practitioners (NP) on validation set B (n=963). The number of cases per condition are presented in Supplementary Table 6. The rightmost columns indicate the average sensitivity for the 26 conditions. Error bars indicate 95% confidence intervals (see Statistical Analysis).

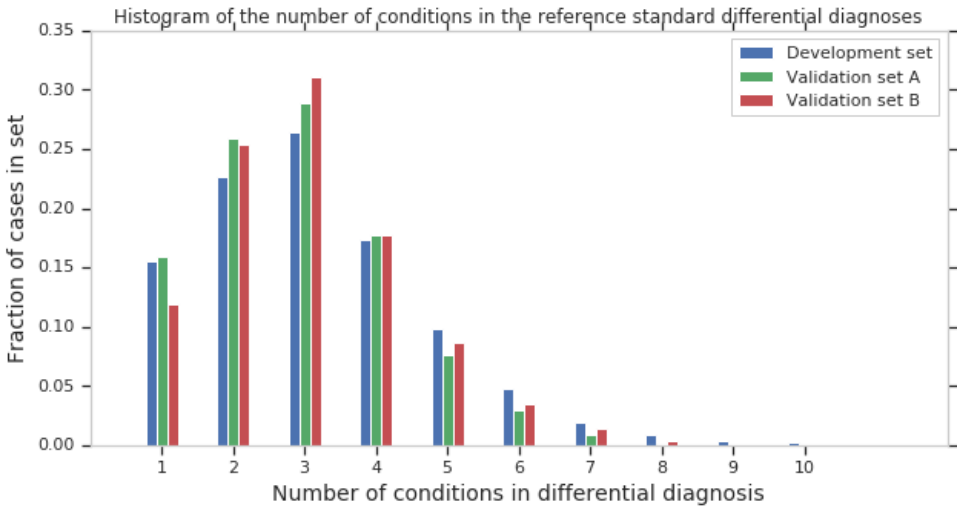


Supplementary Fig. 6 | Performance of the deep learning system (DLS) and clinicians in cases where all three dermatologists determining the reference standard agreed on the primary diagnosis, broken down for each of the 26 categories of skin conditions. a, Top-1

and top-3 sensitivity of the DLS on validation set A (n=3,756). **b**, Top-1 and top-3 sensitivity of the DLS and three types of clinicians: dermatologists (Derm), primary care physicians (PCP), and nurse practitioners (NP) on validation set B (n=963). The number of cases per condition are presented in Supplementary Table 6. The empty bars for the DLS and all clinicians for allergic contact dermatitis are due to the lack of cases that achieved full consensus for that condition. The rightmost columns indicate the average sensitivity for the 26 conditions. Error bars indicate 95% confidence intervals (see Statistical Analysis).

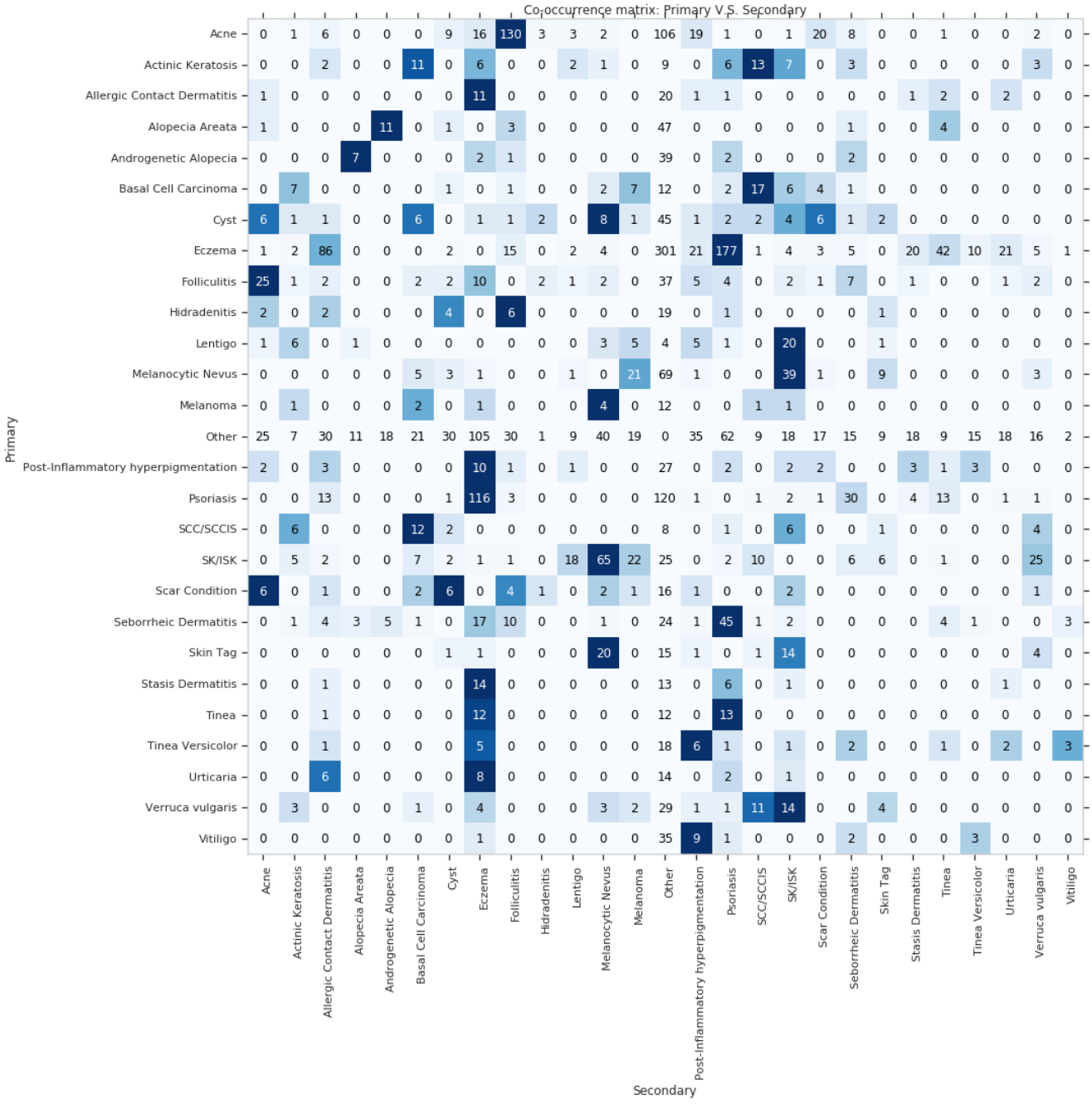


Supplementary Fig. 7 | Labeling tool interface that was designed to present all information that would be available in a teledermatology case. Questions prompts (Supplementary Table 9) are displayed in the left panel, whereas clinical metadata (Supplementary Table 1) are shown in the top right panel and images (up to six per case) are shown in the bottom right panel. Any image could be panned, zoomed, and magnified for closer review. The tool did not enforce any time constraint.

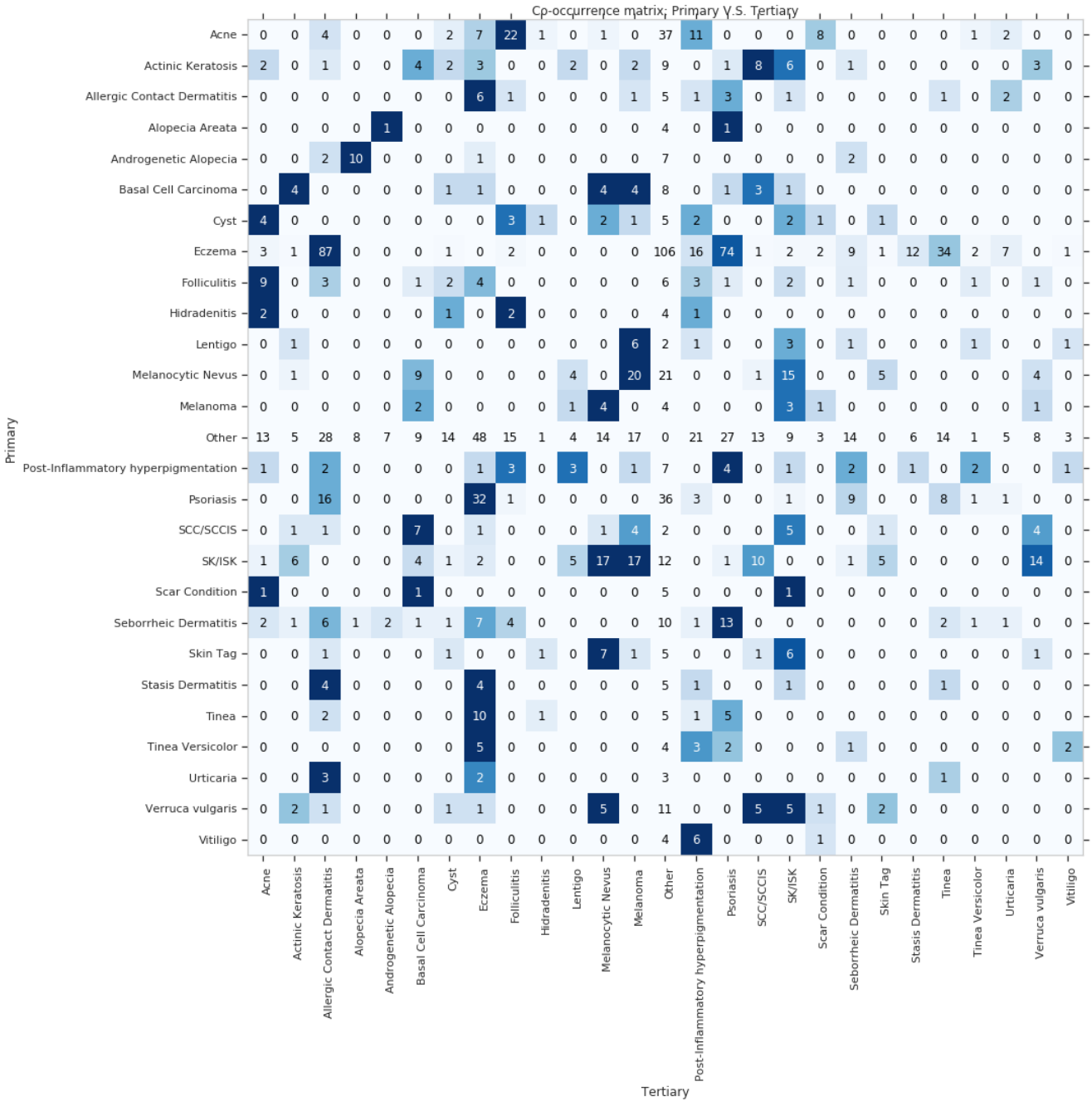


Supplementary Fig. 8 | Histogram of the number of conditions in the reference standard differential diagnoses. Within each set (development set: n=16,114; validation set A: n=3,756; and validation set B: n=963), the differential diagnoses has a 25th percentile length of 2 and a 75th percentile length of 3. The median length was slightly different at 2, 2, 3 for the development set, validation set A, and validation set B, respectively.

a

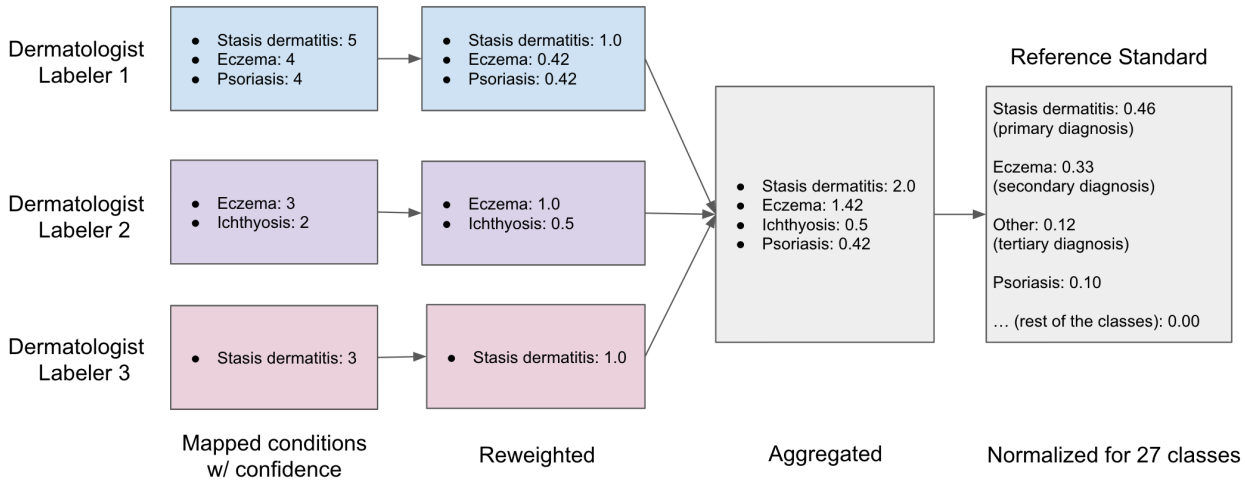

Co-occurrence matrix: Primary V.S. Secondary
Primary
Secondary
Acne 0 1 6 0 0 0 9 16 130 3 3 2 0 106 19 1 0 1 20 8 0 0 1 0 0 2 0
Actinic Keratosis 0 0 2 0 0 11 0 6 0 0 2 1 0 9 0 6 13 7 0 3 0 0 0 0 0 3 0
Allergic Contact Dermatitis 1 0 0 0 0 0 0 11 0 0 0 0 0 20 1 1 0 0 0 0 0 1 2 0 2 0 0
Alopecia Areata 1 0 0 0 11 0 1 0 3 0 0 0 0 47 0 0 0 0 0 1 0 0 4 0 0 0 0
Androgenetic Alopecia 0 0 0 7 0 0 0 2 1 0 0 0 0 39 0 2 0 0 0 2 0 0 0 0 0 0 0
Basal Cell Carcinoma 0 7 0 0 0 0 1 0 1 0 0 2 7 12 0 2 17 6 4 1 0 0 0 0 0 0 0
Cyst 6 1 1 0 0 6 0 1 1 2 0 8 1 45 1 2 2 4 6 1 2 0 0 0 0 0 0
Eczema 1 2 86 0 0 0 2 0 15 0 2 4 0 301 21 177 1 4 3 5 0 20 42 10 21 5 1
Folliculitis 25 1 2 0 0 2 2 10 0 2 1 2 0 37 5 4 0 2 1 7 0 1 0 0 1 2 0
Hidradenitis 2 0 2 0 0 0 4 0 6 0 0 0 0 19 0 1 0 0 0 0 1 0 0 0 0 0 0
Lentigo 1 6 0 1 0 0 0 0 0 0 0 3 5 4 5 1 0 20 0 0 1 0 0 0 0 0 0
Melanocytic Nevus 0 0 0 0 0 5 3 1 0 0 1 0 21 69 1 0 0 39 1 0 9 0 0 0 0 3 0
Melanoma 0 1 0 0 0 2 0 1 0 0 0 4 0 12 0 0 1 1 0 0 0 0 0 0 0 0 0
Other 25 7 30 11 18 21 30 105 30 1 9 40 19 0 35 62 9 18 17 15 9 18 9 15 18 16 2
Post-Inflammatory hyperpigmentation 2 0 3 0 0 0 0 10 1 0 1 0 0 27 0 2 0 2 2 0 0 3 1 3 0 0 0
Psoriasis 0 0 13 0 0 0 1 116 3 0 0 0 0 120 1 0 1 2 1 30 0 4 13 0 1 1 0
SCC/SCCIS 0 6 0 0 0 12 2 0 0 0 0 0 0 8 0 1 0 6 0 0 1 0 0 0 0 4 0
SK/ISK 0 5 2 0 0 7 2 1 1 0 18 65 22 25 0 2 10 0 0 6 6 0 1 0 0 25 0
Scar Condition 6 0 1 0 0 2 6 0 4 1 0 2 1 16 1 0 0 2 0 0 0 0 0 0 0 1 0
Seborrheic Dermatitis 0 1 4 3 5 1 0 17 10 0 0 1 0 24 1 45 1 2 0 0 0 0 4 1 0 0 3
Skin Tag 0 0 0 0 0 0 1 1 0 0 0 20 0 15 1 0 1 14 0 0 0 0 0 0 0 4 0
Stasis Dermatitis 0 0 1 0 0 0 0 14 0 0 0 0 0 13 0 6 0 1 0 0 0 0 0 0 1 0 0
Tinea 0 0 1 0 0 0 0 12 0 0 0 0 0 12 0 13 0 0 0 0 0 0 0 0 0 0 0
Tinea Versicolor 0 0 1 0 0 0 0 5 0 0 0 0 0 18 6 1 0 1 0 2 0 0 1 0 2 0 3
Urticaria 0 0 6 0 0 0 0 8 0 0 0 0 0 14 0 2 0 1 0 0 0 0 0 0 0 0 0
Verruca vulgaris 0 3 0 0 0 1 0 4 0 0 0 3 2 29 1 1 11 14 0 0 4 0 0 0 0 0 0
Vitiligo 0 0 0 0 0 0 0 1 0 0 0 0 0 35 9 1 0 0 0 2 0 0 0 3 0 0 0
Acne
Actinic Keratosis
Allergic Contact Dermatitis
Alopecia Areata
Androgenetic Alopecia
Basal Cell Carcinoma
Cyst
Eczema
Folliculitis
Hidradenitis
Lentigo
Melanocytic Nevus
Melanoma
Other
Post-Inflammatory hyperpigmentation
Psoriasis
SCC/SCCIS
SK/ISK
Scar Condition
Seborrheic Dermatitis
Skin Tag
Stasis Dermatitis
Tinea
Tinea Versicolor
Urticaria
Verruca vulgaris
Vitiligo

b



Supplementary Fig. 9 | Relationship between the primary, secondary, and tertiary diagnoses in the reference standard differential diagnosis in validation set A (n=3,756). a, Co-occurrence matrix representing the secondary diagnosis for each primary diagnosis. **b**, Co-occurrence matrix representing the tertiary diagnosis for each primary diagnosis. Eczema and psoriasis frequently appear together in the differential, and the same applies for other pairs like eczema and tinea, melanocytic nevus and Seborrheic keratosis / irritated seborrheic keratosis

(SK/ISK), and acne and folliculitis. These pairs share visual similarities which can account for their co-occurrence in the same differential diagnosis.



Supplementary Fig. 10 | Illustration of the establishment of reference standard differential diagnosis. In this example, each of the three dermatologists reviewed the case independently and provided a list of diagnoses, each with a confidence value ranging from 1 to 5. Weight for each diagnosis (mapped to the 419421 conditionslist) was determined as the inverse of the rank within each labeler. For the first labeler, since there was a tie between eczema and psoriasis, weights for those were adjusted to evenly distribute between these two ((½ + ⅓) / 2 = 0.42). Answers from different labelers were then aggregated by summing up the weights, before limiting the skin condition classes to 27 and normalizing their weights to sum to 1.

Supplementary Tables

Supplementary Table 1 | Clinical metadata used in this study

Name	Description	Possible values
Self-reported demographic information		
Age	The age of the patient in years, at the time the case was submitted.	A float value ranging from 18 to 90. Values larger than 90 are capped at 90.
Sex	The sex of the patient.	One of: [Female | Male | Other | Unknown]
Race and ethnicity	The race/ethnicity of the patient.	One of: [American Indian or Alaska Native | Asian | Black or African American | Hispanic or Latino | Native Hawaiian or Pacific Islander | White | Neither Hispanic Nor Latino | Not specified | Unknown]
History of the present illness		
Self-reported skin problem	The high level skin problem the patient is seeking help for.	One of: [Acne | Growth or mole | Hair loss | Hair or nail problem | Hair problem | Nail problem | Pigmentary problem | Rash | Other | Unknown]
Symptoms	Any symptoms perceived by the patient.	A list of 8 symptoms (bothersome in appearance, bleeding, increasing in size, darkening, itching, burning, painful, none of the above) with each symptom being one of: [Yes | No | Unknown].
Signs	Any medical signs perceived by the patient.	A list of 7 signs (fever, chills, fatigue, joint pain, mouth sores, shortness of breath, none of the above) with each sign being one of [Yes | No | Unknown].
Duration	The time that the skin problem has persisted.	One of: [One day | Less than one week | One week | Two weeks | One to four weeks | One month | One to three months | Three months | Three to twelve months | Six months | One year | More than one year | More than five years | Since childhood | Since birth | Unknown]
Frequency	Frequency of occurrence of the skin problem.	One of: [Always present | Comes and goes | Unknown]
Past medical history		

Personal history	Personal medical history.	A list of four aspects of the personal history (skin cancer, melanoma, eczema, psoriasis) with each being one of [Yes | No | Unknown].
Family history	Family medical history.	A list of four aspects of the family history (skin cancer, melanoma, eczema, psoriasis) with each being one of [Yes | No | Unknown].
Patient state		
Allergy	Medications the patient is allergic to.	A list of 6 allergies (penicillin, cephalosporin, sulfa, tetracycline, aspirin, other) with each being one of [Yes | No | Unknown].
Drug	If the patient is currently taking any medications.	One of [Yes | No | Unknown].
Pregnancy	If the patient is pregnant.	One of [Yes | No | Unknown].
Nursing	If the patient is nursing.	One of [Yes | No | Unknown].
Medical problem	Whether the patient currently has any medical problems.	One of [Yes | No | Unknown].
Previous treatment state		
Follow-up case	If this is a follow up case.	One of [Yes | No | Unknown].
Biopsy	If there has been a previous biopsy.	One of [Yes | No | Unknown].
Past medication	Whether the patient used medications for the skin problem.	A list of two past medications (prescription drugs, over the counter drugs) with each being one of [Yes | No | Unknown].
Patient adhered to treatments	Whether the patient is following the treatment If the patient received treatment before.	One of: [No | Partially | Yes | Unknown]
Condition after treatments	Progression of the skin problem If the patient received treatment before.	One of: [Improved | Not changed | Worsened | Unknown]

Supplementary Table 2 | Performance of the deep learning system (DLS) and different types of clinicians, on validation set A and validation set B. The reference standard differential diagnoses for each case was determined by the votes of a panel of three board-certified dermatologists. Performance was measured by the agreement of the top-1 and top-3 diagnoses with the primary diagnosis of the panel. The average overlap (AO) directly compares the DLS or clinician-provided ranked differential diagnoses with the panel's full differential diagnoses. The AO ranges from 0 to 1, with higher values indicating better agreement. Numbers in square braces indicate 95% confidence intervals (see Statistical Analysis). Bold indicates the highest value within each column for validation set B.

Dataset	Grader	Top-1		Top-3		Average Overlap (AO)
		Accuracy	Average Sensitivity	Accuracy	Average Sensitivity	
Validation set A (n=3,756)	DLS	0.71 [0.69, 0.72]	0.58 [0.56, 0.60]	0.93 [0.92, 0.94]	0.83 [0.81, 0.85]	0.67 [0.67, 0.68]
Validation set B (enriched subset of set A, n=963)	DLS	**0.66 [0.64, 0.69]**	**0.56 [0.54, 0.59]**	**0.90 [0.88, 0.92]**	**0.82 [0.79, 0.84]**	**0.63 [0.62, 0.65]**
	Derm	0.63 [0.60, 0.65]	0.51 [0.49, 0.54]	0.75 [0.72, 0.77]	0.64 [0.61, 0.66]	0.58 [0.56, 0.59]
	PCP	0.44 [0.42, 0.47]	0.35 [0.33, 0.38]	0.60 [0.58, 0.62]	0.49 [0.47, 0.52]	0.46 [0.44, 0.47]
	NP	0.40 [0.38, 0.43]	0.32 [0.30, 0.34]	0.55 [0.53, 0.58]	0.45 [0.42, 0.47]	0.43 [0.41, 0.44]

Supplementary Table 3 | Performance of the deep learning system (DLS), sliced by self-reported demographic information (including age, sex, race and ethnicity), and Fitzpatrick skin type on validation set A (n=3,756). Metrics used are identical to the ones in Supplementary Table 2. Numbers in square braces indicate 95% confidence intervals (see Statistical Analysis).

Break down	Category	Top-1		Top-3		Average Overlap (AO)
		Accuracy	Average Sensitivity	Accuracy	Average Sensitivity	
Age	[18, 30) (29.5%)	0.76 [0.73, 0.78]	0.55 [0.52, 0.62]	0.95 [0.93, 0.96]	0.80 [0.76, 0.87]	0.71 [0.69, 0.72]
	[30, 40) (19.9%)	0.70 [0.66, 0.73]	0.51 [0.47, 0.58]	0.93 [0.91, 0.94]	0.79 [0.73, 0.84]	0.67 [0.65, 0.69]
	[40, 50) (17.3%)	0.70 [0.66, 0.73]	0.59 [0.54, 0.65]	0.93 [0.91, 0.94]	0.84 [0.80, 0.88]	0.67 [0.65, 0.69]
	[50, 60) (18.6%)	0.68 [0.65, 0.72]	0.61 [0.54, 0.65]	0.92 [0.90, 0.94]	0.81 [0.76, 0.85]	0.66 [0.64, 0.68]
	[60, 90] (14.6%)	0.66 [0.62, 0.70]	0.47 [0.41, 0.53]	0.93 [0.91, 0.95]	0.80 [0.74, 0.85]	0.65 [0.63, 0.67]
Sex	Female (63.1%)	0.71 [0.69, 0.73]	0.58 [0.55, 0.61]	0.93 [0.92, 0.94]	0.83 [0.81, 0.86]	0.67 [0.66, 0.68]
	Male (36.9%)	0.71 [0.68, 0.73]	0.60 [0.56, 0.64]	0.93 [0.91, 0.94]	0.83 [0.80, 0.86]	0.68 [0.66, 0.69]
Race and ethnicity	American Indian or Alaska Native (1.1%)	0.64 [0.50, 0.79]	0.54** [0.41, 0.73]	0.93 [0.86, 1.00]	0.92** [0.81, 1.00]	0.68 [0.61, 0.75]
	Asian (12.6%)	0.75 [0.71, 0.79]	0.56 [0.49, 0.65]	0.95 [0.93, 0.97]	0.85 [0.79, 0.90]	0.68 [0.66, 0.70]
	Black or African American (6.1%)	0.70 [0.63, 0.75]	0.54 [0.46, 0.66]	0.95 [0.92, 0.97]	0.78 [0.74, 0.90]	0.69 [0.65, 0.72]
	Hispanic or Latino (43.4%)	0.71 [0.69, 0.73]	0.55 [0.51, 0.59]	0.93 [0.92, 0.94]	0.81 [0.78, 0.85]	0.68 [0.67, 0.69]
	Native Hawaiian or Pacific Islander (1.6%)	0.70 [0.59, 0.82]	0.58 [0.42, 0.69]	0.97 [0.92, 1.00]	0.78 [0.64, 0.88]	0.65 [0.59, 0.72]

	White (31.3%)	0.69 [0.66, 0.72]	0.60 [0.55, 0.63]	0.92 [0.90, 0.93]	0.81 [0.77, 0.84]	0.67 [0.66, 0.68]
	Not specified (3.9%)	0.70 [0.63, 0.77]	0.57 [0.49, 0.68]	0.93 [0.88, 0.97]	0.83 [0.74, 0.92]	0.67 [0.63, 0.70]
Fitzpatrick skin type	Type I (0.2%)	0.44 [0.11, 0.78]	0.58** [0.25, 1.00]	0.78 [0.44, 1.00]	0.83** [0.67, 1.00]	0.50 [0.40, 0.62]
	Type II (10.2%)	0.71 [0.66, 0.75]	0.61 [0.54, 0.69]	0.91 [0.88, 0.94]	0.79 [0.74, 0.85]	0.66 [0.63, 0.69]
	Type III (64.2%)	0.71 [0.69, 0.73]	0.60 [0.57, 0.62]	0.94 [0.93, 0.95]	0.85 [0.83, 0.87]	0.68 [0.67, 0.69]
	Type IV (19.3%)	0.70 [0.67, 0.73]	0.51 [0.45, 0.57]	0.93 [0.91, 0.94]	0.76 [0.70, 0.82]	0.68 [0.66, 0.70]
	Type V (2.7%)	0.74 [0.65, 0.83]	0.59** [0.48, 0.75]	0.95 [0.90, 0.99]	0.83** [0.78, 0.98]	0.69 [0.65, 0.74]
	Type VI (0.0%)	1.00*	1.00* **	1.00*	1.00* **	0.75*
	Unknown (3.4%)	0.65 [0.55, 0.74]	0.50 [0.42, 0.65]	0.94 [0.89, 0.98]	0.83 [0.76, 0.94]	0.63 [0.58, 0.68]

* : There was only 1 case labeled as Type VI, so confidence intervals were not meaningful.
**: At least ten of the 26 conditions were absent from this subanalysis, resulting in an ill-defined sensitivity for those conditions and an unreliable estimate for average sensitivity.

Supplementary Table 4 | Performance of the deep learning system (DLS) and different types of clinicians on the 419-way classification, on validation sets A and B. The average sensitivity metric was not computed because not all 419 categories had significant representation; missing or rare conditions would skew the numbers. In validation set A for example, 204 unique skin conditions were present as primary diagnoses, and 321 conditions were present as any diagnosis (e.g. secondary, tertiary, etc). Numbers in square braces indicate 95% confidence intervals. Bold indicates the highest value within each column for validation set B.

Dataset	Grader	Top-1 accuracy	Top-3 accuracy	Average Overlap (AO)
Validation set A (n=3,756)	DLS	0.67 [0.66, 0.69]	0.86 [0.85, 0.87]	0.61 [0.60, 0.62]
Validation set B (enriched subset of set A, n=963)	DLS	**0.64 [0.61, 0.67]**	**0.84 [0.82, 0.86]**	**0.57 [0.55, 0.59]**
	Derm	0.61 [0.58, 0.63]	0.72 [0.70, 0.75]	0.52 [0.51, 0.54]
	PCP	0.42 [0.40, 0.45]	0.56 [0.53, 0.58]	0.39 [0.38, 0.41]
	NP	0.40 [0.37, 0.42]	0.51 [0.49, 0.54]	0.37 [0.35, 0.38]

Supplementary Table 5 | Distribution of conditions post biopsy. Abbreviations per Table 2.

Data set	Malignancy (%)	Basal cell carcinoma (%)	Melanoma (%)	SCC/SCCIS (%)
Validation set A (n=3,756)	52 (100%)	32 (61.5%)	6 (11.5%)	14 (26.9%)
Validation set B (n=963)	37 (100%)	19 (51.3%)	5 (13.5%)	13 (35.1%)

Supplementary Table 6 | Number of cases per category of skin condition, filtered by different levels of agreement on the primary diagnosis among dermatologists determining the reference standard.

Condition name	Validation set A			Validation set B		
	No. of cases	No. of cases with agreement by ≥2 dermatologists (%)	No. of cases with agreement by all 3 dermatologists (%)	No. of cases	No. of cases with agreement by ≥2 dermatologists (%)	No. of cases with agreement by all 3 dermatologists (%)
Acne	428	381 (89.0%)	267 (62.4%)	47	36 (76.6%)	21 (44.7%)
Actinic Keratosis	62	40 (64.5%)	19 (30.6%)	43	25 (58.1%)	10 (23.3%)
Allergic Contact Dermatitis	49	27 (55.1%)	2 (4.1%)	36	15 (41.7%)	0 (0.0%)
Alopecia Areata	98	90 (91.8%)	73 (74.5%)	39	35 (89.7%)	31 (79.5%)
Androgenetic Alopecia	56	46 (82.1%)	23 (41.1%)	36	29 (80.6%)	16 (44.4%)
Basal Cell Carcinoma	48	43 (89.6%)	27 (56.2%)	31	27 (87.1%)	17 (54.8%)
Cyst	97	80 (82.5%)	47 (48.5%)	37	29 (78.4%)	16 (43.2%)
Eczema	719	565 (78.6%)	229 (31.8%)	71	38 (53.5%)	11 (15.5%)
Folliculitis	111	64 (57.7%)	20 (18.0%)	43	21 (48.8%)	6 (14.0%)
Hidradenitis	47	41 (87.2%)	31 (66.0%)	37	32 (86.5%)	23 (62.2%)
Lentigo	37	25 (67.6%)	12 (32.4%)	36	29 (74.4%)	12 (33.3%)
Melanocytic Nevus	194	168 (86.6%)	96 (49.5%)	39	29 (72.5%)	16 (41.0%)
Melanoma	27	15 (55.6%)	4 (14.8%)	22	13 (59.1%)	4 (18.2%)
Post Inflammatory Hyperpigmentation	66	37 (56.1%)	8 (12.1%)	38	19 (50.0%)	4 (10.5%)
Psoriasis	365	316 (86.6%)	199 (54.5%)	49	32 (65.3%)	22 (44.9%)
SCC/SCCIS	39	39 (100.0%)	12 (30.8%)	35	33 (94.3%)	12 (34.3%)
SK/ISK	224	203 (90.6%)	118 (52.7%)	44	39 (88.6%)	22 (50.0%)
Scar Condition	69	55 (79.7%)	35 (50.7%)	38	29 (76.3%)	20 (52.6%)
Seborrheic Dermatitis	112	82 (73.2%)	31 (27.7%)	43	27 (62.8%)	11 (25.6%)
Skin Tag	73	68 (93.2%)	39 (53.4%)	35	34 (97.1%)	20 (57.1%)
Stasis Dermatitis	30	18 (60.0%)	6 (20.0%)	29	17 (58.6%)	6 (20.7%)
Tinea	38	27 (71.1%)	14 (36.8%)	35	26 (74.3%)	13 (37.1%)
Tinea Versicolor	37	31 (83.8%)	17 (45.9%)	36	30 (83.3%)	16 (44.4%)
Urticaria	39	28 (71.8%)	14 (35.9%)	38	27 (71.1%)	13 (34.2%)
Verruca vulgaris	88	82 (93.2%)	53 (60.2%)	38	33 (86.8%)	22 (57.9%)
Vitiligo	78	70 (89.7%)	51 (65.4%)	38	34 (89.5%)	25 (65.8%)
Other	910	844 (92.7%)	389 (42.7%)	139	107 (77.0%)	32 (23.0%)

Supplementary Table 7 | Top-1 and top-3 sensitivity averaged across all the skin conditions categories, and with different exclusions on validation set B (n=963). Allergic Contact Dermatitis (ACD) and Post-inflammatory Hyperpigmentation (PIH) are included in this analysis because of the low sensitivity for these conditions by both the deep learning system (DLS) and the three types of clinicians (dermatologists, Derms; primary care physicians, PCPs; and nurse practitioners, NPs). Bold indicates the highest value within each row and each evaluation metric.

Conditions included in the average	Average Top-1 Sensitivity				Average Top-3 Sensitivity			
	DLS	Derm	PCP	NP	DLS	Derm	PCP	NP
All 27 conditions	**0.56**	0.51	0.35	0.32	**0.82**	0.64	0.49	0.45
26 conditions (excludes “Other”)	**0.56**	0.51	0.35	0.32	**0.82**	0.64	0.49	0.45
25 conditions (excludes ACD and PIH)	**0.59**	0.55	0.38	0.35	**0.84**	0.68	0.54	0.49
24 conditions (excludes ACD, PIH, and “Other”)	**0.60**	0.55	0.38	0.34	**0.84**	0.68	0.53	0.48

Supplementary Table 8 | Top-1 and top-3 diagnostic accuracy for the three types of clinicians (dermatologists, Derm; primary care physicians, PCP; and nurse practitioners, NP) on validation set B (n=963). Each clinician graded approximately one-third of the cases (number of cases graded: median = 321, range 320-322). For each clinician, performance of the deep learning system (DLS) is also reported on the same cases graded by that clinician (shaded in gray). Bold indicates the higher of the two: clinician or DLS based on each evaluation metric. In particular, $\text{Accuracy}_{\text{any}}$ measures the agreement of the top-1 and top-3 diagnoses with *any* of the panel of three dermatologists comprising the reference standard.

Clinician Type / ID	Top 1				Top 3				Average Overlap (AO)	
	Accuracy		$\text{Accuracy}_{\text{any}}$		Accuracy		$\text{Accuracy}_{\text{any}}$			
	Clinician	DLS	Clinician	DLS	Clinician	DLS	Clinician	DLS	Clinician	DLS
Derm 1	0.57	**0.68**	0.70	**0.79**	0.68	**0.93**	0.79	**0.98**	0.54	**0.65**
Derm 2	0.66	**0.64**	**0.83**	0.75	0.80	**0.87**	0.93	**0.95**	**0.62**	0.61
Derm 3	0.66	**0.67**	0.81	**0.83**	0.78	**0.91**	0.91	**0.98**	0.59	**0.64**
Derm 4	0.58	**0.66**	0.74	**0.80**	0.69	**0.92**	0.82	**0.97**	0.54	**0.63**
Derm 5	0.64	**0.67**	0.74	**0.79**	0.80	**0.90**	0.88	**0.95**	0.60	**0.62**
Derm 6	0.63	**0.66**	0.75	**0.79**	0.73	**0.88**	0.84	**0.97**	0.56	**0.64**
PCP 1	0.44	**0.70**	0.55	**0.82**	0.67	**0.93**	0.80	**0.97**	0.46	**0.66**
PCP 2	0.49	**0.66**	0.64	**0.77**	0.74	**0.87**	0.86	**0.95**	0.54	**0.61**
PCP 3	0.43	**0.64**	0.61	**0.79**	0.53	**0.90**	0.70	**0.98**	0.43	**0.62**
PCP 4	0.48	**0.66**	0.62	**0.78**	0.50	**0.90**	0.63	**0.96**	0.42	**0.63**
PCP 5	0.43	**0.65**	0.57	**0.78**	0.51	**0.90**	0.65	**0.97**	0.43	**0.63**
PCP 6	0.38	**0.68**	0.53	**0.82**	0.65	**0.90**	0.81	**0.97**	0.47	**0.64**
NP 1	0.42	**0.70**	0.57	**0.81**	0.53	**0.93**	0.66	**0.97**	0.43	**0.65**
NP 2	0.35	**0.63**	0.46	**0.74**	0.55	**0.87**	0.72	**0.95**	0.42	**0.61**
NP 3	0.43	**0.67**	0.56	**0.83**	0.58	**0.91**	0.73	**0.98**	0.44	**0.64**
NP 4	0.38	**0.64**	0.54	**0.76**	0.40	**0.90**	0.56	**0.95**	0.36	**0.63**

NP 5	0.41	**0.71**	0.50	**0.83**	0.60	**0.93**	0.72	**0.98**	0.43	**0.65**
NP 6	0.43	**0.64**	0.59	**0.78**	0.65	**0.87**	0.81	**0.97**	0.47	**0.62**

Supplementary Table 9 | Labeling tool prompts and instructions.

Question		**Possible answers (underlined), with explanations if applicable**
Are multiple conditions present in this case?		Yes*: if more than one condition related to this patient's chief complaint is present Possibly*: if more than one condition may be present No: if there is a single skin condition
Can you describe a differential given the case?		Yes: if one can provide a diagnosis. No*: if one cannot provide any diagnosis. This can be due to poor image quality, minimum pathology, insufficient medical information, etc.
Please provide your top three differential diagnosis:	What is the condition?	SNOMED texts synonyms: an autocomplete menu that contains all synonyms for SNOMED entries pertaining to cutaneous disease is available to select from. If there are several variations of the condition, use the most specific condition that applies to the case. If none found, then: Free text: an additional text field is provided for labelers to enter any free-form text.
	Confidence of diagnosis	5: most certain about the condition. 4: 3: 2: 1: least certain about the condition.

* If these answers are selected, the remaining questions are skipped.

Supplementary Table 10 | Full list of 419 skin conditions that answers from dermatologists, PCPs, and NPs were mapped to. The top 26 conditions on which the DLS was trained and evaluated on are highlighted in bold. The remaining 393 conditions (in aggregate comprising roughly 20% of the cases in this dataset) were mapped to "Other".

A-C	D-H	I-M	N-P	R-Z
Abscess	Deep fungal infection	Ichthyosis	Nail dystrophy due to trauma	RMSF - Rocky Mountain spotted fever
Acanthoma fissuratum	Dental fistula	Idiopathic exfoliative cheilitis	Nasal polyp	Radiation dermatitis
Acanthosis nigricans	Dermatitis herpetiformis	Idiopathic guttate hypomelanosis	Nasolabial dyssebacia	Raynaud's phenomenon
Accessory nipple	Dermatofibroma	IgA pemphigus	Necrobiosis lipoidica	Relapsing polychondritis
Acne	Dermatofibrosarcoma protuberans	Impetigo	Necrolytic acral erythema	Remove from labeling tool
Acne keloidalis	Dermatomyositis	Incontinentia pigmenti	Necrotizing fasciitis	Retention hyperkeratosis
Acquired digital fibrokeratoma	Dermatosis caused by lice	Induced hypopigmentation	Neuralgia paresthetica	Reticular erythematous mucinosis
Acral keratosis	Dermoid cyst of skin	Infected eczema	Neutrophilic eccrine hidradenitis	Reticulate erythematous mucinosis
Acral peeling skin syndrome	Desmoplastic trichoepithelioma	Infected skin ulcer	Nevus anemicus	Reticulohistiocytosis
Acrocyanosis	Diabetic dermopathy	Inflammatory linear verrucous epidermal nevus	Nevus comedonicus	Rheumatoid nodule
Acrodermatitis atrophicans chronica	Diabetic ulcer	Inflicted skin lesions	Nevus depigmentosus	Rhytides
Acropustulosis of infancy	Digital Myxoid Cyst	Ingrown hair	Nevus lipomatosus cutaneous superficialis	Rosacea
Actinic Keratosis	Digital mucous cyst	Injection site disorder	Nevus of Ito	**SCC/SCCIS**
Actinic granuloma	Dissecting cellulitis of scalp	Insect Bite	Nevus of Ota	SJS/TEN
Acute generalised exanthematous pustulosis	Dowling-degos syndrome	Interstitial granulomatous dermatitis	Nevus sebaceous	**SK/ISK**
Adnexal neoplasm	Drug Rash	Intertrigo	Nevus spilus	Scabies
Adult onset still disease	Eccrine carcinoma of skin	Inverted follicular keratosis	Nodular vasculitis	**Scar Condition**
Albinism	Ecthyma	Irritant Contact Dermatitis	Non-melanin pigmentation due to exogenous substance (disorder)	Scleredema
Allergic Contact Dermatitis	Ecthyma gangrenosum	Juvenile xanthogranuloma	Notalgia paresthetica	Sclerodactyly
Alopecia Areata	**Eczema**	Kaposi's sarcoma of skin	O/E - ecchymoses present	Sebaceous adenoma of skin
Alopecia mucinosa	Edema bulla	Keratoderma	Ochronosis	Sebaceous carcinoma
Alopecia neurotica	Epidermal nevus	Keratolysis exfoliativa	Onychocryptosis	Sebaceous hyperplasia
Amyloidosis of skin	Epidermolysis bullosa	Keratosis pilaris	Onychogryphosis	**Seborrheic Dermatitis**
Anagen effluvium	Erosive pustular dermatosis	Knuckle pads	Oncholysis	**Skin Tag**
Androgenetic Alopecia	Eruptive xanthoma	**Lentigo**	Onychomadesis	Skin and soft tissue atypical mycobacterial infection
Anetoderma	Erysipelas	Leprosy	Onychomalacia	Skin atrophy
Angina bullosa hemorrhagica	Erythema ab igne	Leukemia cutis	Onychomatricoma	Skin changes due to malnutrition
Angioedema	Erythema annulare centrifugum	Leukocytoclastic Vasculitis	Onychomycosis	Skin lesion in drug addict
Angiofibroma	Erythema dyschromicum perstans	Leukonychia	Onychopapilloma	Skin striae
Angiokeratoma of skin	Erythema elevatum diutinum	Leukoplakia of skin	Onychorrhexis	Small plaque parapsoriasis
Angiolymphoid hyperplasia with eosinophilia	Erythema gyratum repens	Lichen Simplex Chronicus	Onychoschizia	Small vessel thrombosis of skin
Angiosarcoma of skin	Erythema marginatum	Lichen nitidus	Oral fibroma	Sneddon-Wilkinson disease
Animal bite - wound	Erythema migrans	Lichen planopilaris	Osteoarthritis	**Stasis Dermatitis**
Apocrine cystadenoma	Erythema multiforme	Lichen planus/lichenoid eruption	Osteoma	Subungual fibroma
Arsenical keratosis	Erythema nodosum	Lichen sclerosus	Osteoma cutis	Sweet syndrome
Arterial ulcer	Erythrasma	Lichen spinulosus	Otitis externa	Symmetrical dyschromatosis of extremities
Arteriovenous malformation	Erythromelalgia	Lichen striatus	Paget disease	Syphilis
Atrophic glossitis	Erythromelanosis follicularis faciei et colli	Lichenoid keratosis	Palisaded neutrophilic granulomatous dermatitis	TMEP - telangiectasia macularis eruptiva perstans
Atrophoderma	Fat necrosis	Lichenoid myxedema	Palmar pit	Tattoo
Atrophoderma vermiculatum	Fibrofolliculoma	Linear IgA disease	Papilloma of skin	Telangiectasia disorder
Atypical Nevus	Flagellate erythema	Lipoatrophy	Parapsoriasis	Telogen effluvium
Atypical fibroxanthoma of skin	Flegels disease	Lipodermatosclerosis	Paronychia	Thrombophlebitis
B-Cell Cutaneous Lymphoma	Flushing	Lipoid proteinosis	Pearly penile papules	**Tinea**
Basal Cell Carcinoma	Focal epithelial hyperplasia of skin	Lipoma	Pemphigoid gestationis	**Tinea Versicolor**
Beau's lines	**Folliculitis**	Lipschütz ulcer	Pemphigus foliaceus	Torus palatinus
Becker's nevus	Folliculitis decalvans	Livedo reticularis	Pemphigus paraneoplastica	Trachyonychia
Benign neoplasm of nail apparatus	Fordyce spots	Livedoid vasculopathy	Pemphigus vulgaris	Traction alopecia
Benign neural tumor	Foreign body	Lobomycosis	Perforating dermatosis	Traumatic bulla
Benign salivary gland tumor	Foreign body reaction of the skin	Local infection of wound	Perichondritis of auricle	Traumatic ulcer
Blistering distal dactylitis	Fox-Fordyce disease	Longitudinal melanonychia	Perioral Dermatitis	Triangular alopecia
Blue sacral spot	Frontal fibrosing alopecia	Lymphadenopathy	Periungual fibroma	Trichostasis spinulosa
Bowenoid papulosis	Ganglion cyst	Lymphangioma	Perleche	Trichotillomania
	Geographic tongue	Lymphedema	Phimosis	Trigeminal trophic syndrome
	Giant cell tumor	Lymphomatoid papulosis	Photodermatitis	
		Madarosis	Phrynoderma	
		Malignant cylindroma	Piezogenic pedal papule	

Brachioradial pruritus	Glomus tumour of skin	Malignant eccrine spiradenoma	Pigmented fungiform papillae	Tripe palms
Breast cancer	Gout	Mastocytoma	Pigmented purpuric eruption	Tuberculosis of skin and subcutaneous tissue
Bullosis diabeticorum	Graft versus host disease	Mastocytosis	Pilomatricoma	Ulceration in Behcet disease
Bullous Pemphigoid	Granular parakeratosis	Median rhomboid glossitis	Pilonidal cyst	**Urticaria**
Burn of skin	Granuloma annulare	Melanin pigmentation due to exogenous substance	Pincer nail deformity	Urticaria multiforme
Bursitis	Granuloma faciale	**Melanocytic Nevus**	Pinkus tumor	Varicose veins of lower extremity
Cafe au lait macule	Granulomatous cheilitis	**Melanoma**	Pitted keratolysis	Venous Stasis Ulcer
Calcinosis cutis	Grover's disease	Melanotic macule	Pityriasis alba	**Verruca vulgaris**
Calciphylaxis cutis	Hailey Hailey disease	Melasma	Pityriasis amiantacea	Viral Exanthem
Candida	Hair nevus	Merkel Cell Carcinoma	Pityriasis lichenoides	**Vitiligo**
Canker sore	Hair sinus	Microcystic adnexal carcinoma	Pityriasis rosea	Warty dyskeratoma
Carotene pigmentation of skin	Hairy tongue	Milia	Pityriasis rotunda	Wells' syndrome
Cellulitis	Half-and-half nail	Miliaria	Pityriasis rubra pilaris	Wooly hair
Central centrifugal cicatricial alopecia	Hand foot and mouth disease	Molluscum Contagiosum	Pleomorphic fibroma	Xanthoma
Chancroid	Head lice	Morphea/Scleroderma	Poikiloderma	Xerosis
Chemical leukoderma	Hemangioma	Morsicatio buccarum	Porokeratosis	Yellow nail syndrome
Chicken pox exanthem	Hematoma of skin	Mucocele	Porphyria cutanea tarda	Zoon's balanitis
Chilblain	Hemorrhoid	Mucocutaneous venous malformation	**Post-Inflammatory hyperpigmentation**	Zosteriform reticulate hyperpigmentation
Chondrodermatitis nodularis	Hemosiderin pigmentation of skin		Post-Inflammatory hypopigmentation	
Cicatricial Pemphigoid	Herpes Simplex		Pressure ulcer	
Clavus	Herpes Zoster		Pressure-induced dermatosis	
Clear cell acanthoma	**Hidradenitis**		Pretibial myxedema	
Clubbing of fingers	Hirsutism		Primary cutaneous sarcoma	
Collagenoma	Hordeolum internum		Progressive macular hypomelanosis	
Colloid milium	Hyperhidrosis		Prurigo nodularis	
Comedone	Hypersensitivity		Pruritic urticarial papules and plaques of pregnancy	
Condyloma acuminatum	Hypertrichosis		Pseudocyst of auricle	
Confluent and reticulate papillomatosis			Pseudolymphoma	
Congenital alopecia			Pseudopelade	
Connective tissue nevus			**Psoriasis**	
Crohn disease of skin			Psychogenic alopecia	
Cutaneous T Cell Lymphoma			Pterygium of nail	
Cutaneous capillary malformation			Puncture wound - injury	
Cutaneous collagenous vasculopathy			Purpura	
Cutaneous larva migrans			Pyoderma Gangrenosum	
Cutaneous leishmaniasis			Pyogenic granuloma	
Cutaneous lupus				
Cutaneous lymphadenoma				
Cutaneous metastasis				
Cutaneous myiasis				
Cutaneous neurofibroma				
Cutaneous neuroma				
Cutaneous sarcoidosis				
Cutaneous schistosomiasis				
Cutaneous sporotrichosis				
Cutis laxa				
Cutis verticis gyrata				
Cylindroma of skin				
Cyst				

Supplementary Table 11 | Hyperparameters for training the deep learning system.

Image augmentations	Image size: 459×459 pixels Saturation delta: [0.5597, 1.2749] Contrast delta: [0.9997, 1.7705] Brightness max delta: 0.1148 Hue max delta: 0.0251 Rotation: [-150, 150] (degrees) Flipping: horizontal, vertical
Bounding box augmentations	Minimum overlap with any pathologic region: 0.2 Aspect ratio: [0.9, 1.1] Proportion over the original image: [0.05, 1.0]
Metadata augmentations	Dropout rate: 0.1
Learning rate schedule (exponential decay schedule)	Base rate: 0.001 Decay rate: 0.99 Number of epochs per decay: 2.0
Adam optimizer	Decay for the first moment estimates: 0.9 Decay for the second moment estimates: 0.999 Epsilon: 0.1
Batch size	8
Regularization	Prelogits dropout rate: 0.2 Weight decay: 0.00004 Batch norm decay: 0.9997
Loss function	Softmax cross-entropy with class-specific weights
Class weighting	Weight for each class is determined as: $1 \,/\, c^{1-s}$ Where c is the class counts over the training set, and s is a smoothing factor of 0.7.

Supplementary Table 12 | Performance of the deep learning system (DLS) and different types of clinicians, on validation sets A and B. This is similar to Supplementary Table 2, except performance was measured by the agreement of the top-1 and top-3 diagnoses with *any* of the panel of three dermatologists comprising the reference standard. In other words, whether the top k predictions of the DLS or clinician captures the primary diagnosis of any member of the panel. For agreement with a differential diagnosis based on the "votes" of the panel, see Supplementary Table 2. Numbers in square braces indicate 95% confidence intervals (see Statistical Analysis). Bold indicates the highest value within each column for validation set B.

Dataset	"Grader"	Top-1	Top-3
		$\text{Accuracy}_{\text{any}}$	$\text{Accuracy}_{\text{any}}$
Validation set A (n=3,756)	DLS	0.82 [0.81, 0.83]	0.98 [0.97, 0.98]
Validation set B (enriched subset of set A, n=963)	DLS	**0.79 [0.77, 0.82]**	**0.97 [0.96, 0.98]**
	Derm	0.76 [0.74, 0.78]	0.86 [0.84, 0.88]
	PCP	0.59 [0.56, 0.61]	0.74 [0.72, 0.76]
	NP	0.54 [0.51, 0.56]	0.70 [0.68, 0.72]

Supplementary Table 13 | Performance of the deep learning system (DLS), stratified by self-reported demographic information (including age, sex, race and ethnicity), and Fitzpatrick skin type on validation set A (n=3,756). Metrics used are identical to the ones in Supplementary Table 12. Numbers in square braces indicate 95% confidence intervals (see Statistical Analysis).

Breakdown	Category	Top-1	Top-3
		$Accuracy_{any}$	$Accuracy_{any}$
Age	[18, 30) (29.5%)	0.85 [0.83, 0.87]	0.98 [0.97, 0.99]
	[30, 40) (19.9%)	0.81 [0.78, 0.83]	0.97 [0.96, 0.98]
	[40, 50) (17.3%)	0.83 [0.80, 0.86]	0.98 [0.98, 0.99]
	[50, 60) (18.6%)	0.81 [0.79, 0.84]	0.97 [0.96, 0.98]
	[60, 90] (14.6%)	0.78 [0.75, 0.82]	0.98 [0.96, 0.99]
Sex	Female (63.1%)	0.83 [0.82, 0.85]	0.98 [0.97, 0.98]
	Male (36.9%)	0.81 [0.79, 0.83]	0.97 [0.96, 0.98]
Race and ethnicity	American Indian or Alaska Native (1.1%)	0.76 [0.64, 0.880]	0.95 [0.88, 1.00]
	Asian (12.6%)	0.85 [0.82, 0.88]	0.98 [0.97, 0.99]
	Black or African American (6.1%)	0.82 [0.77, 0.87]	0.98 [0.96, 1.00]
	Hispanic or Latino (43.4%)	0.82 [0.81, 0.84]	0.98 [0.97, 0.98]
	Native Hawaiin or Pacific Islander (1.6%)	0.77 [0.66, 0.87]	1.00 [1.00, 1.00]
	White (31.3%)	0.81 [0.79, 0.83]	0.97 [0.97, 0.98]
	Not specified (3.9%)	0.83 [0.78, 0.89]	0.99 [0.97, 1.00]
Fitzpatrick skin type	Type I (0.2%)	0.78 [0.56, 1.00]	0.89 [0.67, 1.00]

	Type II (10.2%)	0.83 [0.79, 0.87]	0.97 [0.96, 0.99]
	Type III (64.2%)	0.82 [0.81, 0.84]	0.98 [0.97, 0.99]
	Type IV (19.3%)	0.82 [0.79, 0.85]	0.97[0.96, 0.98]
	Type V (2.7%)	0.84 [0.77, 0.91]	0.98 [0.95, 1.00]
	Type VI (0.0%)	1.00*	1.00*
	Unknown (3.4%)	0.76 [0.68, 0.84]	0.97 [0.93, 1.00]

* : There was only 1 case labeled as Type VI, so confidence intervals were not meaningful.