Supplementary Online Content

Jain A, Way D, Gupta V, et al. Development and assessment of an artificial intelligence–based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. *JAMA Netw Open*. 2021;4(4):e217249. doi:10.1001/jamanetworkopen.2021.7249

eMethods. Procedures and Metrics

eFigure 1. Detailed User Interface of the Artificial Intelligence (AI)–Based Assistive Tool

eFigure 2. States Where Readers Are Licensed to Practice

eFigure 3. Summary of Prior Clinical Experience of Primary Care Physician (PCP) and Nurse Practitioner (NP) Readers

eFigure 4. Diagnostic Accuracy of Primary Care Physicians (PCPs) and Nurse Practitioners (NPs)

eFigure 5. Sensitivity for the Top 26 Most Common Skin Conditions, With vs Without Artificial Intelligence (AI) Assistance

eFigure 6. Impact of Artificial Intelligence (AI) Assistance Stratified by the Position of the Correct Diagnosis in the Assistant’s Interface

eFigure 7. Impact of Artificial Intelligence (AI) Assistance Stratified by the AI’s Confidence Level (as Indicated by 5 Dots in the Assistant’s Interface)

eFigure 8. Impact of Case Difficulty (as Measured by Interdermatologist Agreement Within the Panel Providing the Reference Diagnosis) on the Performance of Unassisted and Assisted Readers

eFigure 9. Changes During the Course of the Study for Top-1 Agreement, Top-3 Agreement, Average Overlap, and Median Diagnosis Time

eFigure 10. Impact of Artificial Intelligence (AI) Assistance Stratified by Estimated Fitzpatrick Skin Type for Top-1 Agreement, Top-3 Agreement, and Average Overlap

eFigure 11. Changes in Top-3 Diagnostic Accuracy by Primary Care Physicians (PCPs) and Nurse Practitioners (NPs) Without Artificial Intelligence (AI) Assistance vs When Using AI Assistance on Nonreferred Cases and Referred Cases

eFigure 12. Breakdown of Accuracy of Cases Based on Confidence of Primary Care Physicians (PCPs) and Nurse Practitioners (NPs)

eFigure 13. Sample Cases Evaluated Against Reference Diagnoses Provided by a Panel of 3 Dermatologists

eFigure 14. Sample Cases Evaluated Against Biopsy-Provided Diagnoses

eFigure 15. Reader Perceptions of the Value of Training Materials and Labeling Tool

eFigure 16. Reader Perceptions of the Value of Artificial Intelligence (AI) Assistance

eFigure 17. Perceived Usefulness of Different Features in the Artificial Intelligence (AI)-Based Assistance for Primary Care Physician (PCP) and Nurse Practitioner (NP) Readers

eTable 1. Clinical Metadata Used in This Study for Both AI Algorithm Input and Presented to Clinicians

eTable 2. Improvement in Top-1 Agreement and Average Overlap Between 2 Dermatologist Panels Using the Aggregated ("Voting")-Based Approach

eTable 3. Additional Metrics (Top-1 and Top-3 Agreement, Average Overlap, and Kappa) for Measuring Agreement Rates of PCPs and NPs, Respectively, With the Dermatologist Panel

eTable 4. Subgroup Analysis for 3 Categories of Skin Condition for Growths, Erythematousquamous and Papulosquamous Skin Disease, and Hair Loss

eTable 5. Rates of Recommending a Biopsy and Referral for PCPs and NPs

© 2021 Jain A et al. *JAMA Network Open.*
Sample Size Estimation

To estimate the sample size, number of cases, and number of clinicians needed, we first conducted a “feasibility study” using an independent dataset (ie, the “tune set” from prior work), and with a different set of 6 clinicians (4 PCPs, 2 NPs). We then used bootstrapping to simulate the effect of different sample sizes and numbers of clinicians. Our simulations suggested that having 1,048 cases read by 20 readers would yield statistical power exceeding 90% even if the effect size was 25% of that observed in the feasibility study. Therefore, we settled on 20 PCPs and 20 NPs.

Alternative Statistical Analyses

We also conducted a permutation test that preserved the structure of the reading cohorts in the study. In this test, the assignment of each reader (PCPs or NPs) to RC-1 and RC-2 was permuted in each trial. This assignment determined the modality for each case review. The corresponding confidence intervals were computed by first resampling readers within each cohort and then resampling cases. Using these methods, the effect sizes were nearly identical; the top-1 agreement with reference diagnosis improved with AI assistance by 10% (95%CI 8-12%, p<0.001) for PCPs and 12% (95%CI 10-15%, p<0.001) for NPs.

In addition to the permutation test, we used a generalized linear mixed effects model. AI assistance was the fixed effect and random effects were raters, cases, the interaction between AI assistance and raters, and the interaction between AI assistance and cases. This model was fitted with the lme4 package in R via: glmer("agreement ~ ai_assistance + (1+ai_assistance|case_id) + (1+ai_assistance|rater_id)"), which estimates the coefficients by maximum likelihood method. The Wald test showed significant effects of AI assistance for both PCPs and NPs, with effect sizes of 0.68 (p<0.001) and 0.84 (p<0.001) on the logit scale. These correspond to odds ratios of 1.98 (95%CI 1.77-2.21) and 2.31 (95%CI 1.98-2.70), respectively.

A simplified model excluding the interaction between AI assistance and the case showed similar results: effect sizes of 0.64 (p<0.001) and 0.80 (p<0.001) on the logit scale. These correspond to odds ratios of 1.89 (95%CI 1.71-2.08) and 2.23 (95%CI 1.95-2.54), respectively.

Reader Characteristics

Readers were recruited via a medical staffing company (Vituity, Emeryville, CA), with instructions to optimize for broad geographical diversity, and without further intervention from the study investigators. Initially, 20 PCPs and 20 NPs were recruited and began the study. However, 4 readers (2 PCPs and 2 NPs) were unavailable to complete the study (study completion rate: 0%, 0%, 3%, 13%), and were
replaced by an additional 4 readers recruited by the staffing company. The readers’ reasons were personal and unrelated to the study. We next describe the final cohort of 40 readers.

The 20 PCPs were all US board certified, had an average of 11.3 years of experience (range 2-32), and were located across 12 US states (eFigures 2-3). The 20 NPs had an average of 13.1 years of experience (range 2-34), and were practicing in primary care settings without physician supervision across 9 US states. Post-randomization to reading cohorts, the average years of experience were 11.3 (range 2-23) for PCPs in RC-1, 11.3 (range 2-32) for PCPs in RC-2, 13.3 (range 2-26) for NPs in RC-1, and 12.8 (range 6-34) for NPs in RC-2.

For additional exploratory analysis, 2 U.S. board-certified dermatologists (4 and 7 years of experience post-residency) read cases following the same procedures as the PCPs and NPs. These 2 dermatologists had participated in our prior study of comparing stand-alone AI with clinicians (ie, without AI assistance) 12 months before the commencement of this study.

**AI Tool Interface**

The AI tool provided several pieces of information for each predicted skin condition. First, the predicted likelihood was displayed as a 5-point “Assistance Confidence”, where prediction likelihoods were rounded down to the nearest half-point. Next, additional pre-compiled dermatologist-provided reference information was presented for each suggested skin condition: (1) a description of the skin condition; (2) the suggested clinical workup for the type of condition; (3) treatment information; (4) clinical pearls such as risk factors and other information relevant to the condition; (5) differential diagnoses including other entities the condition is commonly mistaken for and how to distinguish them from the present condition; and (6) a set of textbook images of the present skin condition. The pre-compiled reference information provided for each skin condition was the same for every case in which the condition was suspected. Finally, this reference information was augmented with visually similar images from cases previously diagnosed as the suspected skin conditions. These images were obtained by comparing AI’s internal numerical representation of the case under examination (“embeddings”) to the numerical representations of each case in the library of cases used to develop the algorithm (See “Similar Image Retrieval” section).

**Onboarding Process**

All clinicians in this study participated in a self-led training (“onboarding”) process to familiarize themselves with the AI assistant tool. This process involved watching an introductory tutorial video that walked through the tool, and reviewing a digital slide deck that provided an overview of the development and evaluation of the underlying deep learning system: the cases it was developed on, the reference standard label used to train and evaluate it. The slides also covered aspects that the AI has not been validated to handle: poor image quality, minimal skin pathology, multiple skin conditions, and cases for which rendering a confident diagnosis not possible by dermatologists. The interface itself was also described in detail, including each item in Figure 1 and indications for when the predicted condition was not present in large quantities in the original training dataset (labeled as “Based on low data”). The instructions additionally covered data entry and how to use the “search as you type”
interface to provide their differential diagnosis and instructed the readers to provide as specific a diagnosis as possible. Finally, two sample cases (independent of the 1,048 study cases) were provided to improve familiarization with the tool.

Additional Evaluation Metrics

Two additional evaluation metrics (in addition to agreement rate with the differential diagnoses) were used to assess concordance between the dermatologists and NPs or PCPs respectively.

The first metric, “top-3 agreement” measures if any of the top three differential diagnoses by the PCP or NP matched the reference diagnosis. The second metric, “average overlap” mathematically quantifies the similarity between two differential diagnoses. The metric ranges from 0 to 1, where values close to 1 indicate that the differential diagnosis of the primary care provider is relatively complete compared to the dermatologist-provided reference differential diagnosis (Eng et al., British Journal of Dermatology, 2020). Results from this analysis are presented in eTable 3 and are qualitatively similar to those from the primary analysis.

Similar Image Retrieval

For each case, several possible skin condition predictions were displayed by the AI, as described in the Methods. For each of these skin conditions, a “similar images panel” (eFigure 1) was constructed by retrieving images from cases of the skin condition.

Specifically, the retrieved cases were from the “development set” used to develop the algorithm. Matching images had similar AI-based numerical representations (“embeddings”) to those of the case being reviewed, where similarity between images was quantified using the L2-distance between the PreLogit embeddings for each image. Because each case contained multiple images, we computed a case similarity score each pair of cases. Case similarity was the minimum distance between any pair of images across the two cases (i.e. minimum pairwise distance). Images from the 10 most similar cases were included in the “similar images” panel.

Impact of AI Accuracy on Assistance

In several pre-planned subanalyses, we examined the relationship between algorithm performance and reader diagnostic accuracy. First, we examined effects associated with the rank of the correct prediction within the list of predictions (eFigure 6). In this analysis, when the reference diagnosis was the top condition (1st position) predicted by the AI algorithm, assistance was associated with large increases in agreement with reference diagnoses for NPs and PCPs. However, when the reference diagnosis was in the 2nd-5th AI-ranked predictions, assistance was associated with little change or downward trends in top-1 agreement for both NPs and PCPs. For the 12% of cases where the reference diagnosis did not appear in the (up to 5) predictions, decreases in top-1 agreement with the reference diagnoses were observed. When considering the top-3 agreement however, the trends were different, with a sustained trend of improved agreement with the reference diagnoses even when the

© 2021 Jain A et al. JAMA Network Open.
reference diagnosis was in the second or third position. These nuances highlight how the different evaluation metrics are complementary by providing different perspectives of the AI assistance.

In terms of the full differential, we found that there was a consistent increase in the average overlap for both PCPs and NPs when the AI itself had an average overlap above 0.3 (eFigure 6E). Conversely, when the AI had a low average overlap (below 0.1), there was a consistent negative impact on both PCPs and NPs. This is particularly important when considering the ability of the AI to improve the readers’ overall differential diagnosis instead of their ability to capture only the single reference diagnosis.

Next, we examined the impact of assistance on sensitivity for each of 26 skin conditions (eFigure 5). Top-1 agreement with the reference diagnoses trended upwards with assistance for 25 of 26 conditions, including notable increases for skin conditions eczema, melanocytic nevus, SK/ISK, acne, seborrheic dermatitis, verruca vulgaris, scar condition, basal cell carcinoma, urticaria, tinea versicolor, lentigo, and androgenetic alopecia. These were also the conditions for which the AI algorithm had a high standalone top-1 agreement. Conversely, for the one condition where assistance was associated with decreased reader diagnostic accuracy, melanoma, the AI’s top-1 agreement was low at below 15%, albeit with wide confidence intervals (it was the rarest condition amongst the 26).

These subanalyses highlight the importance of algorithm accuracy in AI assistance and the need to have measures in place to mitigate over-reliance on the assistant, particularly when predictions are incorrect. These results are consistent with those from Tschandl et al. (Nature Medicine, 2020), who found that, by deliberately presenting incorrect diagnoses to readers, their system was able to mislead clinicians with various amounts of expertise. Other studies have also highlighted negative impacts of overreliance (Alberdi et al. Academic Radiology, 2004) and have explored techniques to mitigate these effects through AI explainability (Cai et al. ACM HCI Conference, 2019).

In our study, our interface presented multiple skin condition predictions for each case, displayed AI confidence and reference information alongside predictions, and alerted users when predicted skin conditions were under-represented in algorithm development. However, future work is needed to further understand and mitigate over-reliance; we hypothesize this work could ultimately integrate improvements to the assistant interface, additional reader training with the AI tools, and improved underlying algorithm accuracy.

Sensitivity Analysis to Potential Technical Issues

We discovered after the conclusion of the study that a technical issue occurred during a 25-hour period (1pm PDT, March 18, 2020 to 2pm PDT March 19, 2020). This issue may have made it more difficult for the clinicians to select the intended answer for the 424 cases (1.0% of the full study) that were reviewed during that time. However, the visible selected answer was correctly logged, and there were no reports of case reviewing difficulty from any of the clinicians. To ensure that our conclusions were unaffected, we repeated the main analyses with all reviews from this time period excluded. Minimal changes were seen in this sensitivity analysis, with a 1% or small change (if any) in the metrics: for PCPs top-1 accuracy were 48% without assistance and 58% with assistance (delta of 10%, 95%CI 8-

© 2021 Jain A et al. JAMA Network Open.
11, p<0.001), and for NPs top-1 accuracy were 45% without assistance and 58% with assistance (delta of 12%, 95%CI 10-14, p<0.001).
eFigure 1. Detailed User Interface of the Artificial Intelligence (AI)–Based Assistive Tool

The left pane contains the case, which is comprised of one or more photographs of the skin condition, along with basic demographic information such as age and self-reported sex and the past medical history. In the assisted mode, the right pane shows up to 5 of the AI’s top skin condition predictions. The AI’s confidence in each prediction is shown as colored dots, and additional information such as sample images from an atlas are displayed as well. In the non-assisted mode, this pane is missing.

© 2021 Jain A et al. JAMA Network Open.
eFigure 2. States where readers are licensed to practice.

For PCPs these were: Arizona, California, Florida, Georgia, Illinois, Iowa, Kentucky, Michigan, North Carolina, Oregon, Texas, and Washington. For NPs, the states were: Arizona, Colorado, Georgia, Idaho, Maryland, Montana, Vermont, Washington, Wyoming. The dermatologists were from: California, Colorado, Hawaii, Iowa, Massachusetts, New Jersey, New York, South Carolina, Tennessee and Texas.
**eFigure 3.** Summary of Prior Clinical Experience of Primary Care Physician (PCP) and Nurse Practitioner (NP) Readers
eFigure 4. Diagnostic Accuracy of Primary Care Physicians (PCPs) and Nurse Practitioners (NPs)

Results without artificial intelligence (AI) assistance, versus when using AI assistance measured using several metrics. Every clinician provided their differential diagnosis (several rank-ordered conditions), which were then mapped to 419 skin conditions. (A) Top-1 agreement measures how often their primary diagnosis agrees with a panel of dermatologists; while (B) top-3 agreement measures how often at least one of the top 3 diagnoses provided by the PCP or NP matched the panel of dermatologists. (C) The average overlap (AO) metric evaluates how well the full differential provided by each clinician matches the differential provided by the panel of dermatologists. (D) The agreement with ≥1 dermatologist measures the top-1 agreement with at least one dermatologist’s top reference diagnosis (instead of the aggregated reference diagnosis from voting).
eFigure 5. Sensitivity for the Top 26 Most Common Skin Conditions, With vs Without Artificial Intelligence (AI) Assistance

The green lines represent the AI algorithm agreement with reference diagnoses, and the green shading represents the corresponding 95% confidence intervals. The top-3 sensitivity reflects the sensitivity obtained by considering the top 3 differential diagnoses from the clinicians, and is substantially higher.
than the top-1 sensitivity. Abbreviations used: SCC / SCCIS=Squamous Cell Carcinoma / Squamous Cell Carcinoma In Situ; SK / ISK=Seborrheic Keratosis / Irritated Seborrheic Keratosis.
eFigure 6. Impact of Artificial Intelligence (AI) Assistance Stratified by the Position of the Correct Diagnosis in the Assistant’s Interface
"N/A" indicates that the AI assistant did not display the correct diagnosis in any position. (A) Histogram illustrating the proportion of cases for which the correct primary diagnosis was in each position in the AI assistant's display. (B, C, D) Top-1 agreement, top-3 agreement, and average overlap (AO) of the unassisted and assisted clinicians based on the position of the correct diagnosis. (E) The AO of the clinicians as a function of the AI assistant's AO.
eFigure 7. Impact of Artificial Intelligence (AI) Assistance Stratified by the AI’s Confidence Level (as Indicated by 5 Dots in the Assistant’s Interface)

(A) Histogram illustrating the proportion of cases for which the AI displayed that level of confidence. The level of confidence is displayed in half-point increments on the interface. For ease of visualization, half point values have been rounded down. (B, C, D) Top-1 agreement, top-3 agreement, and average overlap (AO) of the unassisted and assisted clinicians based on the confidence indicated by the AI assistant.

© 2021 Jain A et al. JAMA Network Open.
**eFigure 8.** Impact of Case Difficulty (as Measured by Interdermatologist Agreement Within the Panel Providing the Reference Diagnosis) on the Performance of Unassisted and Assisted Readers

![Bar chart showing agreement in top-1 within DERM panel for PCP, NP, and DERM for unanimous, majority, and no agreement cases.]
(A) top-1 agreement, (B) top-3 agreement, (C) average overlap, and (D) median diagnosis time. Each batch contains 50 cases, with the exception of the last 2 batches (which contained 24 cases each). The assistance modality (ie, assisted vs. unassisted) alternated for each reader between batches (Figure 1). With assistance, the median time to diagnose increased from 89s to 94s for PCPs, and from 77s to 84s for NPs. Because the review time decreased dramatically during the first 4 batches (presumably as the readers gained familiarity with the interface and task), we computed the median review time after excluding these batches. The deltas remained; assistance increased the median time from 82s to 88s for PCPs and from 72s to 79s for NPs.
eFigure 10. Impact of Artificial Intelligence (AI) Assistance Stratified by Estimated Fitzpatrick Skin Type for Top-1 Agreement, Top-3 Agreement, and Average Overlap

(A) top-1 agreement, (B) top-3 agreement, and (C) average overlap.
**eFigure 11.** Changes in Top-3 Diagnostic Accuracy by Primary Care Physicians (PCPs) and Nurse Practitioners (NPs) Without Artificial Intelligence (AI) Assistance vs When Using AI Assistance on Nonreferred Cases and Referred Cases

(A) non-referred cases, and (B) referred cases.
**eFigure 12.** Breakdown of Accuracy of Cases Based on Confidence of Primary Care Physicians (PCPs) and Nurse Practitioners (NPs)

| Confidence of top diagnosis | PCP  | NP  | DERM |
|-----------------------------|------|-----|------|
| > 90%                       | 0.64 | 0.58| 0.79 |
| 75-90%                      | 0.47 | 0.38| 0.54 |
| 50-75%                      | 0.35 | 0.32| 0.48 |
| 25-50%                      | 0.32 | 0.26| 0.35 |
| < 25%                       | 0.27 | 0.16| 0.32 |

**Mean top-1 agreement**

|        | UNASSISTED | UNASSISTED | UNASSISTED |
|--------|------------|------------|------------|
|        |            |            |            |
|        | 0.0        | 0.2        | 0.4        |
|        | 0.6        | 0.8        | 1.0        |

© 2021 Jain A et al. *JAMA Network Open*.
eFigure 13. Sample Cases Evaluated Against Reference Diagnoses Provided by a Panel of 3 Dermatologists

(A-B) Sample cases where AI assistance was associated with increased reader agreement with reference diagnoses. (C) Sample case where AI assistance was associated with decreased agreement with the reference diagnosis.
eFigure 14. Sample Cases Evaluated Against Biopsy-Provided Diagnoses

(A-B) Sample cases where AI assistance was associated with increased reader agreement with biopsy-provided diagnoses. (C) Sample case where AI assistance was associated with decreased agreement with the biopsy-provided diagnosis.
**eFigure 15. Reader Perceptions of the Value of Training Materials and Labeling Tool**

Error bars indicate mean ± 95% confidence interval response, if treating 5-point Likert scale as an interval measure. Over 80% of PCPs and NPs in the study reported afterward that they understood the AI feedback moderately well or better. Only 4 readers across both backgrounds reported "Somewhat" understanding the AI assistance; none replied "Not at all".

© 2021 Jain A et al. *JAMA Network Open.*
eFigure 16. Reader Perceptions of the Value of Artificial Intelligence (AI) Assistance

Error bars indicate mean ± 95% confidence interval response, if treating 5-point Likert scale as an interval measure.
eFigure 17. Perceived Usefulness of Different Features in the Artificial Intelligence (AI)-Based Assistance for Primary Care Physician (PCP) and Nurse Practitioner (NP) Readers

Colored lines next to each feature name correspond to the outline of the feature on the example screenshot. n=20 for PCPs and n=19 for NPs; one NP did not respond to the survey. There was a range of perceived usefulness of different features, with matching conditions and textbook images being rated useful most often; and clinical pearls and workup least often.
The AI algorithm interprets clinical photographs of skin conditions and patient medical history, as described below, in making its classification. The same information was available to clinicians in making their diagnoses.

| Metadata category       | 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 | Past medical history | Family history | Patient state | Previous treatment state |
|--------------------------------|-----------|----------------------|----------------|---------------|--------------------------|
| Frequency of occurrence of the skin problem. | One of: [Always present | Comes and goes | Unknown] | One of: [Yes | No | Unknown]. | A list of four aspects of the personal history (skin cancer, melanoma, eczema, psoriasis) with each being one of [Yes | No | Unknown]. |
| 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]. | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| 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]. | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| 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]. | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| If the patient is currently taking any medications. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| If the patient is pregnant. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| If the patient is nursing. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| Whether the patient currently has any medical problems. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| If this is a follow up case. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| If there has been a previous biopsy. | | | | One of: [Yes | No | Unknown]. | One of: [Yes | No | Unknown]. |
| Whether the patient used medications for the skin problem. | | | | One of: [Yes | No | Unknown]. | A list of two past medications (prescription drugs, over the counter drugs) with each being one of [Yes | No | Unknown]. |
| Whether the patient is following the treatment if the patient received treatment before. | | | | One of: [No | Partially | Yes | Unknown] | One of: [No | Partially | Yes | Unknown] |
| Progression of the skin problem if the patient received treatment before. | | | | One of: [Improved | Not changed | Worsened | Unknown] | One of: [Improved | Not changed | Worsened | Unknown] |
Data from previous work (Liu et al. Nature Medicine, 2020). The agreement rate between the dermatologist panel (reviewing the clinical images and history) with the biopsy results (n=152 cases) is 0.55.

| Metric           | Between 2 randomly-selected dermatologists (from 2 independent panels) | Between 2 panels |
|------------------|-------------------------------------------------------------------------|------------------|
| Top-1 agreement  | 0.61                                                                    | 0.74             |
| Average overlap  | 0.54                                                                    | 0.63             |
eTable 3. Additional Metrics (Top-1 and Top-3 Agreement, Average Overlap, and Kappa) for Measuring Agreement Rates of PCPs and NPs, Respectively, With the Dermatologist Panel

Top-3 agreement measures if any of the top three differential diagnoses by the PCP or NP matched the reference diagnosis. Average overlap mathematically quantifies the similarity between differential diagnoses provided by the dermatologist panels and PCPs and NPs respectively. Because the “expected” accuracy in Kappa is noisy when certain classes are rare, kappa was computed across 27 categories: the top 26 most common conditions, with the remaining conditions aggregated as an “other” category, consistent with prior work.

|                | Top-1 Agreement | Top-3 Agreement | Average overlap | Kappa          |
|----------------|-----------------|-----------------|-----------------|----------------|
|                | Unassisted (%)  | Assisted (%)    | Differen (95%CI)| Unassisted (%) | Assisted (%)    | Differen (95%CI) | Unassisted (%) | Assisted (%)    | Differen (95%CI) | Unassisted (%) | Assisted (%)    | Differen (95%CI) |
| PCPs           | 48              | 58              | 10 (8-11)       | 58             | 68             | 10 (9-12)       | 42             | 49             | 7 (6-8)         | 0.501           | 0.597           | 0.096 (0.077-0.114) |
| NPs            | 46              | 58              | 12 (10-14)      | 54             | 66             | 12 (11-14)      | 40             | 49             | 9 (7-10)        | 0.468           | 0.595           | 0.127 (0.108-0.146) |
eTable 4. Subgroup Analysis for 3 Categories of Skin Condition for Growths, Erythematosquamous and Papulosquamous Skin Disease, and Hair Loss

(A) presents the specific conditions present in this study and included in the subanalysis for each category; (B,C) present the impact on top-1 and top-3 agreement with AI assistance. Bold indicates the higher value between unassisted and assisted columns. Note that the growth analysis here is based on the reference diagnosis from the dermatologist panel, which provides a larger sample sizes compared to solely the cases with available biopsy results.

| Category       | Subcategory   | Conditions in the subcategory                                                                 | No. of cases |
|----------------|---------------|---------------------------------------------------------------------------------------------|--------------|
| Growth         | Malignant     | Angiosarcoma of skin, Atypical fibroxanthoma of skin, B-Cell Cutaneous Lymphoma, Basal Cell Carcinoma, Atypical Nevus, Becker's Nevus | 89           |
|                | Precancerous  | Actinic Keratosis, Arsenical keratosis, Bowenoid papulosis                                    | 41           |
|                | Benign        | Acanthoma fissuratum, Accessory nipple, Acquired digital fibrokeratoma, Acral keratosis, Adnexal neoplasm, Angiokeratoma of skin, Apocrine cystadenoma, Atypical Nevus, Becker's Nevus, Benign neoplasm of nail apparatus, Benign neural tumor, Benign salivary gland tumor, Blue sacral spot, Cafe au lait macule, Chondrodermatitis nodularis, Clavus, Clear cell acanthoma | 325          |

© 2021 Jain A et al. JAMA Network Open.
| Collagenoma,            | Colloid milium,          |
|------------------------|--------------------------|
| Condyloma acuminatum,  | Comedone,                |
| Connective tissue nevus, | Cutaneous lymphadenoma,  |
| Cutaneous neurofibroma, | Cutaneous neuroma,       |
| Cyst,                  | Cylindroma of skin,      |
| Dermatofibroma,        | Ganglioni cyst,          |
| Dermoid cyst of skin,  | Giant cell tumor,        |
| Desmoplastic trichoepithelioma, | Glomus tumour of skin, |
| Digital Myxoid Cyst,    | Granuloma faciale,       |
| Digital mucous cyst,    | Inflammatory linear verrucous epidermal nevus, |
| Epidermal nevus,        | Inverted follicular keratosis, |
| Eruptive xanthoma,      | Juvenile xanthogranuloma,|
| Fibrofolliculoma,       | Knuckle pads,            |
| Focal epithelial hyperplasia of skin, | Lentigo, |
| Ganglion cyst,          | Lichenoid keratosis,     |
| Giant cell tumor,       | Lipoma,                  |
| Glomus tumour of skin,  | Lymphangioma,            |
| Granuloma faciale,      | Mastocytoma,             |
| Inflammatory linear verrucous epidermal nevus, | Melanocytic Nevis, |
| Inverted follicular keratosis, | Melanotic macule, |
| Juvenile xanthogranuloma, | Milia,                |
| Knuckle pads,           | Molluscum Contagiosum,   |
| Lentigo,                | Mucocele,                |
| Lichenoid keratosis,    | Nasal polyp,             |
| Lipoma,                 | Nevus comedonicus,       |
| Lymphangioma,           | Nevus lipomatosus cutaneous superficialis, |
| Mastocytoma,            | Nevus of Ito,            |
| Melanocytic Nevis,      | Nevus of Ota,            |
| Melanotic macule,       | Nevus sebaceous,         |
| Milia,                  | Nevus spilus,            |
| Molluscum Contagiosum,  | Onychomatricoma,         |
| Mucocele,               | Onychopapilloma,         |
| Nasal polyp,            | Oral fibroma,            |
| Nevus comedonicus,      | Osteoma cutis,           |
| Nevus lipomatosus cutaneous superficialis, | Papilloma of skin, |
| Nevus of Ota,           | Pearly penile papules,   |
| Nevus sebaceous,        | Periungual fibroma,      |
| Nevus spilus,           | Pigmented fungiform papillae, |
| Onychomatricoma,        | Pilomatrixicoma,         |
| Onychopapilloma,        | Pilonidal cyst,          |
| Oral fibroma,           | Pleomorphic fibroma,     |
| Osteoma cutis,          | Porokeratosis,           |
| Papilloma of skin,      | Pseudocyst of auricle,   |
| Pearly penile papules,  | Pseudolymphoma,          |
| Periungual fibroma,     | Pyogenic granuloma,      |
| Pigmented fungiform papillae, | Rheumatoid nodule,   |
| Pilomatrixicoma,        | SK/ISK,                  |
| Pilonidal cyst,         |                           |
| Pleomorphic fibroma,    |                           |
| Porokeratosis,          |                           |
| Pseudocyst of auricle,  |                           |
| Pseudolymphoma,         |                           |
| Pyogenic granuloma,     |                           |
| Rheumatoid nodule,      |                           |
| SK/ISK,                 |                           |

© 2021 Jain A et al. JAMA Network Open.
| Scar Condition, Sebaceous adenoma of skin, Sebaceous hyperplasia, Skin Tag, Subungual fibroma, Tattoo, Torus palatinus, Verruca vulgaris, Warty dyskeratoma, Xanthoma  |
|---|---|---|
| Infectious | Syphilis, Tinea, Tinea Versicolor | 63 |
| Erythematosquamous and papulosquamous skin disease | Non-infectious | Allergic Contact Dermatitis, Cutaneous T Cell Lymphoma, Cutaneous lupus, Cutaneous sarcoidosis, Dermatomyositis, Drug Rash, Eczema, Erythema annulare centrifugum, Erythema gyratum repens, Hypersensitivity, Irritant Contact Dermatitis, Lichen planus/lichenoid eruption, Parapsoriasis, Photodermatitis, Pityriasis lichenoides, Pityriasis rosea, Pityriasis rotunda, Pityriasis rubra pilaris, Psoriasis, Radiation dermatitis, Seborrheic Dermatitis, Small plaque parapsoriasis | 200 |
| Hair loss (1) | Scarring | Acne keloidalis, Alopecia mucinosa, Central centrifugal cicatricial alopecia, Dissecting cellulitis of scalp, Folliculitis decalvans, Frontal fibrosing alopecia, Lichen planopilaris, Traction alopecia | 3 |
| Non-scarring | Alopecia Areata, Alopecia aretica, Anagen effluvium, Androgenetic Alopecia, Congenital alopecia, Madarosis, Psychogenic alopecia, Syphilis, Telogen effluvium, Triangular alopecia, Trichotillomania | 72 |
| Hair loss (2) | Androgenetic Alopecia | Androgenetic Alopecia | 34 |
| Alopecia Areata | Alopecia Areata | 37 |

© 2021 Jain A et al. JAMA Network Open.
| Category                     | Subcategory                                      | Top-1 Agreement (%) | Top-1 PPV (%) |
|------------------------------|--------------------------------------------------|---------------------|---------------|
|                              |                                                  | PCP Unassisted     | PCP Assisted  | NP Unassisted | NP Assisted |
|                              |                                                  | 60 59 64 67        | 66 71 55 66  |               |             |
| Growth                       | Malignant                                        | 53 60 36 56        | 60 67 44 67  |               |             |
|                              | Precancerous                                     | 73 77 68 75        | 92 92 93 93  |               |             |
|                              | Benign                                           |                     |               |               |             |
| Erythematous and papulosquamous skin disease | Infectious                                       | 61 74 63 76        | 65 77 63 77  |               |             |
|                              | Non-infectious                                   | 52 59 52 61        | 91 93 89 93  |               |             |
| Hair loss (1)                | Scarring                                         | 13 47 13 40        | 17 52 44 67  |               |             |
|                              | Non-scarring                                     | 82 88 83 91        | 100 100 100 100 |           |             |
| Hair loss (2)                | Androgenetic Alopecia                            | 64 75 58 70        | 91 93 86 90  |               |             |
|                              | Alopecia Areata                                  | 76 82 75 84        | 96 98 82 89  |               |             |

© 2021 Jain A et al. *JAMA Network Open.*
| Category                              | Subcategory                                      | Top-3 Agreement (%) | Top-3 PPV (%) |
|---------------------------------------|-------------------------------------------------|---------------------|---------------|
|                                       |                                                 | PCP Unassisted      | PCP Assisted  | NP Unassisted | NP Assisted |
|                                       |                                                 | 75                  | 75            | 78            | 77          |
|                                       | Malignant                                       | 50                  | 54            | 47            | 51          |
|                                       | Precancerous                                    | 46                  | 52            | 36            | 54          |
|                                       | Benign                                          | 88                  | 89            | 90            | 90          |
| Growth                                | Infectious                                      | 69                  | 83            | 69            | 82          |
|                                       | Non-infectious                                  | 87                  | 88            | 85            | 88          |
|                                       | Erythematous squamous and papulosquamous skin disease | 62                  | 68            | 61            | 70          |
|                                       | Scarring                                        | 17                  | 53            | 20            | 43          |
|                                       | Non-scarring                                    | 87                  | 91            | 86            | 92          |
| Hair loss (1)                         | Androgenetic Alopecia                           | 70                  | 82            | 61            | 74          |
|                                       | Alopecia areata                                 | 82                  | 85            | 80            | 86          |
| Hair loss (2)                         |                                                 | 86                  | 89            | 76            | 79          |
eTable 5. Rates of Recommending a Biopsy and Referral for PCPs and NPs

|       | Rate of recommending biopsy | Rate of recommending referral |
|-------|-----------------------------|-------------------------------|
|       | Unassisted (%) | Assisted (%) | Difference (95% CI) | Unassisted (%) | Assisted (%) | Difference (95% CI) |
| PCPs  | 24.5            | 23.0           | -1.4 (-0.2 - -2.7)  | 42.3            | 39.6           | -2.7 (-1.2 - -4.4)  |
| NPs   | 24.9            | 23.0           | -1.8 (-0.6 - -3.0)  | 35.1            | 32.2           | -2.9 (-1.3 - -4.4)  |

© 2021 Jain A et al. JAMA Network Open.