JAMA Network Open

Original Investigation | Dermatology

# Development and Assessment of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices

Ayush Jain, MS; David Way, ME; Vishakha Gupta, MS; Yi Gao, PhD; Guilherme de Oliveira Marinho, BS; Jay Hartford, MS; Rory Sayres, PhD; Kimberly Kanada, MD; Clara Eng, PhD; Kunal Nagpal, MS; Karen B. DeSalvo, MD, MPH, MSc; Greg S. Corrado, PhD; Lily Peng, MD, PhD; Dale R. Webster, PhD; R. Carter Dunn, MS, MBA; David Coz, MS; Susan J. Huang, MD; Yun Liu, PhD; Peggy Bui, MD, MBA; Yuan Liu, PhD

## Abstract

**IMPORTANCE** Most dermatologic cases are initially evaluated by nondermatologists such as primary care physicians (PCPs) or nurse practitioners (NPs).

**OBJECTIVE** To evaluate an artificial intelligence (AI)–based tool that assists with diagnoses of dermatologic conditions.

**DESIGN, SETTING, AND PARTICIPANTS** This multiple-reader, multiple-case diagnostic study developed an AI-based tool and evaluated its utility. Primary care physicians and NPs retrospectively reviewed an enriched set of cases representing 120 different skin conditions. Randomization was used to ensure each clinician reviewed each case either with or without AI assistance; each clinician alternated between batches of 50 cases in each modality. The reviews occurred from February 21 to April 28, 2020. Data were analyzed from May 26, 2020, to January 27, 2021.

**EXPOSURES** An AI-based assistive tool for interpreting clinical images and associated medical history.

**MAIN OUTCOMES AND MEASURES** The primary analysis evaluated agreement with reference diagnoses provided by a panel of 3 dermatologists for PCPs and NPs. Secondary analyses included diagnostic accuracy for biopsy-confirmed cases, biopsy and referral rates, review time, and diagnostic confidence.

**RESULTS** Forty board-certified clinicians, including 20 PCPs (14 women [70.0%]; mean experience, 11.3 [range, 2-32] years) and 20 NPs (18 women [90.0%]; mean experience, 13.1 [range, 2-34] years) reviewed 1048 retrospective cases (672 female [64.2%]; median age, 43 [interquartile range, 30-56] years; 41 920 total reviews) from a teledermatology practice serving 11 sites and provided 0 to 5 differential diagnoses per case (mean [SD], 1.6 [0.7]). The PCPs were located across 12 states, and the NPs practiced in primary care without physician supervision across 9 states. The NPs had a mean of 13.1 (range, 2-34) years of experience and practiced in primary care without physician supervision across 9 states. Artificial intelligence assistance was significantly associated with higher agreement with reference diagnoses. For PCPs, the increase in diagnostic agreement was 10% (95% CI, 8%-11%; $P < .001$), from 48% to 58%; for NPs, the increase was 12% (95% CI, 10%-14%; $P < .001$), from 46% to 58%. In secondary analyses, agreement with biopsy-obtained diagnosis categories of maglignant, precancerous, or benign increased by 3% (95% CI, −1% to 7%) for PCPs and by 8% (95% CI, 3%-13%) for NPs. Rates of desire for biopsies decreased by 1% (95% CI, 0-3%) for PCPs and 2% (95% CI, 1%-3%) for NPs; the rate of desire for referrals decreased by 3% (95% CI, 1%-4%) for PCPs and NPs. Diagnostic agreement on cases not indicated for a dermatologist referral increased by 10% (95% CI,

(continued)

## Key Points

**Question** Can artificial intelligence help primary care physicians and nurse practitioners diagnose skin conditions more accurately?

**Findings** In this diagnostic study of 20 primary care physicians and 20 nurse practitioners reviewing 1048 retrospective cases, artificial intelligence assistance was significantly associated with higher agreement with diagnoses made by a dermatologist panel, with an increase from 48% to 58% for primary care physicians and an increase from 46% to 58% for nurse practitioners. These outcomes correspond to a benefit for 1 in every 8 to 10 cases.

**Meaning** Artificial intelligence may help clinicians diagnose skin conditions more accurately in primary care practices, where most skin diseases are initially evaluated.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

*Abstract (continued)*

8%-12%) for PCPs and 12% (95% CI, 10%-14%) for NPs, and median review time increased slightly by 5 (95% CI, 0-8) seconds for PCPs and 7 (95% CI, 5-10) seconds for NPs per case.

**CONCLUSIONS AND RELEVANCE** Artificial intelligence assistance was associated with improved diagnoses by PCPs and NPs for 1 in every 8 to 10 cases, indicating potential for improving the quality of dermatologic care.

*JAMA Network Open.* 2021;4(4):e217249. doi:10.1001/jamanetworkopen.2021.7249

## Introduction

With 2 billion people affected globally,[1] skin conditions are a leading cause of morbidity. The examination of some skin conditions by dermatologists results in significantly higher diagnostic accuracy[2-4] and is associated with better clinical outcomes[5] than nondermatologist examination. However, owing to lack of access to dermatologists, only 28% of skin cases are seen by a specialist[6]; therefore, nonspecialists play a pivotal role in the assessment of skin lesions and initiation of clinical management and referrals.[7] The diagnostic accuracy of nonspecialists is reportedly only 24% to 70%,[4,8-10] suggesting that currently available resources, such as dermatology textbooks, medical information portals, and online image search engines, remain insufficient to guide nonspecialists.

Several algorithms incorporating artificial intelligence (AI) have been developed to help interpret both clinical[11-15] and dermoscopic[16-23] images for a variety of skin conditions, and the effect of AI-based support on dermoscopic images has been studied.[15,24] However, an open question remains as to whether AI assistance can help primary care physicians (PCPs) and nurse practitioners (NPs) diagnose skin conditions from clinical images (ie, taken without specialized equipment).

We developed an AI-based tool and conducted a multiple-reader, multiple-case diagnostic study in which PCPs and independently practicing NPs retrospectively reviewed skin cases from a teledermatology service, representing 120 different skin conditions. We used randomization to ensure readers reviewed each case only once, either with or without AI assistance. Our primary objective was to measure the AI assistance–associated changes in diagnostic accuracy of PCPs and NPs without specialist training in dermatology.

## Methods

This study was approved by the Quorum Institutional Review Board, Seattle, Washington, and deemed exempt from informed consent because all data and images were deidentified. The Standards for Reporting of Diagnostic Accuracy (STARD)[25] reporting guideline was followed for this study.

### The AI Tool

Liu et al[26] previously described an AI algorithm that provides a differential diagnosis given clinical photographs of skin conditions and the medical history (eTable 1 in the Supplement). Their AI model was developed using 16 114 cases and used a convolutional neural network to output prediction scores across 419 skin conditions. In the present study, we created a web-based tool using the AI model described by Liu et al by incorporating user experience insights (**Figure 1**).

The tool provides information about the case, including demographic information, history of present illness, and other elements of the patient's medical history. For each case, 1 to 6 images were available for review (median, 4), and readers could toggle between or zoom in on images. Primary care physicians and NPs reviewed these cases using a laptop and could consult additional resources as they would in clinical practice.

The AI assistance component of the web-based tool was only available during the assisted mode of the study (described below). At the top of the panel, the interface displayed the skin conditions that were output by the AI, sorted in order of the AI's predicted likelihood scores. Artificial intelligence predictions with low scores (<0.05) were removed, and the list was limited to 5 skin conditions to avoid presenting extraneous information. Each condition could be clicked on to display additional information (Figure 1 and eFigure 1 and the AI Tool Interface section in the eMethods in the Supplement).

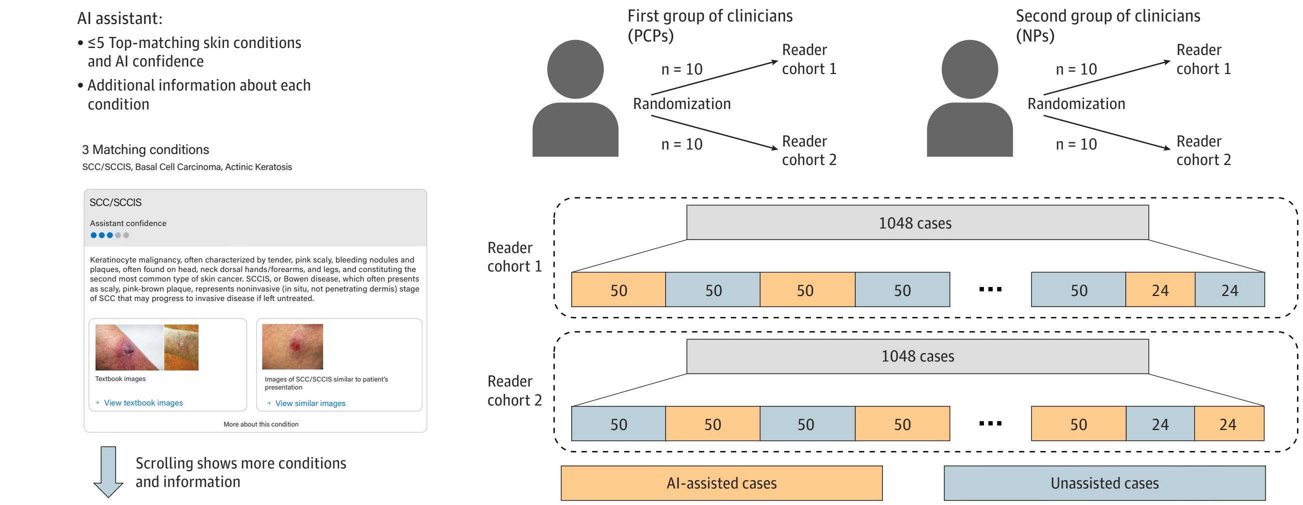
## Study Design

To evaluate whether this tool could assist primary care clinicians in diagnosing skin conditions, we conducted a multiple-reader, multiple-case diagnostic study with 20 PCPs and 20 NPs (Figure 1). The characteristics of the clinicians are described in the Reader Characteristics section of the eMethods and eFigures 2 and 3 in the Supplement. Before reviewing the study cases, each reader was presented with materials describing how to use the AI assistant and given the opportunity to practice using the AI assistant with 2 sample cases (independent of the study cases). Additional details of this training[27] can be found in the Onboarding Process section in the eMethods in the Supplement.

The study used cases from 2 retrospective data sets from California and Hawaii previously used to validate the AI algorithm.[26] Specifically, the prior study used a validation set A and a subset (validation set B) enriched for rarer conditions via random sampling stratified by condition. Validation set B (963 cases) was included in its entirety. From validation set A, all 85 cases for which biopsy results were available were also included to yield a total of 1048 cases (**Table**). None of the PCPs or NPs in this study previously reviewed these cases, and the AI algorithm used was identical to the one used in the previous study.[26]

Each reader was randomly assigned to 1 of 2 reader cohorts. The 2 reader cohorts read the same cases but with the opposite assistance modalities (ie, unassisted vs AI assisted) for each case. To reduce effects associated with switching modalities, the 1048 cases were divided into batches of 50 cases (except the last 48 cases, which were divided into 2 batches of 24 cases), and the assistance modality switched after each batch of cases. For the first batch of 50 cases, reader cohort 1 reviewed

Figure 1. User Interface of the Artificial Intelligence (AI)–Based Assistive Tool and the Study Design



Keratinocyte malignancy, often characterized by tender, pink scaly, bleeding nodules and plaques, often found on head, neck dorsal hands/forearms, and legs, and constituting the second most common type of skin cancer. SCCIS, or Bowen disease, which often presents as scaly, pink-brown plaque, represents noninvasive (in situ, not penetrating dermis) stage of SCC that may progress to invasive disease if left untreated.

The AI assistant shows as many as 5 top predictions of skin conditions, with the confidence in each prediction shown as colored dots and additional information (eg, sample images from an atlas) available with a click. More details are available in eFigure 1 in the Supplement. The study was designed as a multiple-reader, multiple-case (MRMC) study comprising 1048 cases. Two groups of clinicians (primary care physicians [PCPs] and nurse practitioners [NPs]) reviewed each case with or without AI assistance. The modality alternated every 50 cases. For every case, each clinician was instructed to rank as many as 3 differential diagnoses using a search-as-you-type interface and selecting matching skin conditions from a list of 3961 conditions. If their desired skin condition was not present, clinicians could provide free-text entries. All skin conditions were mapped to a list of 419 conditions. SCC indicates squamous cell carcinoma; SCCIS, SCC in situ.

Table. Patient Characteristics in the Original Data Set and Final Enriched Data Set

| Characteristic | Data set<br>Full study (n = 1048)[a] | <br>Cases with diagnoses from histologic findings (n = 152)[b] |
|---|---|---|
| Years | 2017-2018 | 2017-2018 |
| No. of sites | 11 | 10 |
| No. of images included in study | 3935 | 413 |
| No. of patients included in study | 1016 | 152 |
| Age, median (IQR), y[a] | 43 (30-56) | 49 (35-59) |
| Sex, No. (%) | | |
| Female | 672 (64.2) | 99 (65.1) |
| Male | 375 (35.8) | 53 (34.9) |
| Race and ethnicity, No. (%) | | |
| American Indian or Alaska Native | 9 (0.9) | 0 |
| Asian | 102 (9.7) | 5 (3.3) |
| Black or African American | 66 (6.3) | 5 (3.3) |
| Hispanic or Latino | 447 (42.7) | 59 (38.8) |
| Native Hawaiian or Pacific Islander | 20 (1.9) | 2 (1.3) |
| White | 365 (34.9) | 80 (52.6) |
| Not specified | 38 (3.6) | 1 (0.7) |
| Fitzpatrick skin type (6 types), No. (%) | | |
| I | 2 (0.2) | 2 (1.3) |
| II | 109 (10.4) | 17 (11.2) |
| III | 668 (63.8) | 111 (73.0) |
| IV | 205 (19.6) | 14 (9.2) |
| V | 25 (2.4) | 1 (0.7) |
| VI | 0 | 0 |
| Unknown | 38 (3.6) | 7 (4.6) |
| Skin conditions based on primary diagnosis, No. (%)[c] | | |
| Acne | 40 (3.8) | NA |
| Actinic keratosis | 39 (3.7) | 1 (0.7) |
| Allergic contact dermatitis | 25 (2.4) | NA |
| Alopecia areata | 37 (3.5) | NA |
| Androgenetic alopecia | 32 (3.1) | NA |
| Basal cell carcinoma | 36 (3.4) | 32 (21.1) |
| Cyst | 32 (3.1) | 1 (0.7) |
| Eczema | 53 (5.1) | NA |
| Folliculitis | 32 (3.1) | 3 (2.0) |
| Hidradenitis | 34 (3.2) | NA |
| Lentigo | 32 (3.1) | 3 (2.0) |
| Melanocytic nevus | 61 (5.8) | 28 (18.4) |
| Melanoma | 20 (1.9) | 6 (3.9) |
| Postinflammatory hyperpigmentation | 28 (2.7) | NA |
| Psoriasis | 40 (3.8) | NA |
| SCC/SCCIS | 34 (3.2) | 14 (9.2) |
| SK/ISK | 52 (5.0) | 13 (8.6) |
| Scar condition | 34 (3.2) | 2 (1.3) |
| Seborrheic dermatitis | 37 (3.5) | NA |
| Skin tag | 36 (3.4) | 3 (2.0) |
| Stasis dermatitis | 25 (2.4) | NA |
| Tinea | 31 (3.0) | 1 (0.7) |
| Tinea versicolor | 34 (3.2) | NA |
| Urticaria | 33 (3.2) | NA |
| Verruca vulgaris | 37 (3.5) | 8 (5.3) |
| Vitiligo | 36 (3.4) | NA |
| Other[d] | 116 (11.1) | 65 (42.8) |

Abbreviations: IQR, interquartile range; NA, not applicable; SCC/SCCIS, squamous cell carcinoma/squamous cell carcinoma in situ; SK/ISK, seborrheic keratosis/irritated seborrheic keratosis.

[a] One case was removed from the study for logistical reasons.

[b] Of 165 cases, 13 had equivocal biopsy results and were excluded from the biopsy analysis. A total of 141 cases had growths and 53 were malignant.

[c] Enrichment was performed to avoid skew toward common conditions (eg, acne and eczema) as described previously and additionally to include all available cases with biopsy confirmation.

[d] Conditions with fewer than 10 cases each.

these cases with AI assistance, whereas reader cohort 2 reviewed the same cases unassisted. The next batch of cases were reviewed in the opposite modality (Figure 1). By ensuring each reader reviewed each case only once in either the assisted or unassisted modality, this design eliminated any memory effect associated with a crossover study (where memorable cases may inflate the diagnostic performance when reviewed a second time by the same readers).[28,29]

During the case reviews, the readers either provided their top differential diagnoses or indicated that they were unable to diagnose a case. They also answered a few questions on their intended clinical next steps for each case (see the Study End Points section below). Reviews were performed without time constraint. These reviews occurred from February 21 to April 28, 2020.

## Reference Diagnoses

Reference diagnoses were provided by a panel of dermatologists.[26] Briefly, 3 US board-certified dermatologists (from a pool of 12) independently reviewed each case. The dermatologists participated in the study via Advanced Clinical, Deerfield, Illinois; had 5 to 13 years of experience (mean [SD], 7.2 [2.7] years); and practiced in multiple states, including Colorado, Hawaii, Iowa, Maryland, New York, South Carolina, Tennessee, and Texas. Reference diagnoses were obtained using a previously described collective intelligence approach, which results in more reproducible diagnoses than diagnoses obtained by individual dermatologist review (eTable 2 in the Supplement).[26,30] This approach assigns a vote to each diagnosis based on its ranking: the first diagnosis in a dermatologist's differential was given a weight of 1/1 = 1; the secondary diagnosis was given a weight of 1/2 = 0.5. The votes for each diagnosis were summed across the 3 dermatologists, and the top-voted diagnosis was considered the primary diagnosis of the panel.

Agreement was also assessed against biopsy-confirmed diagnoses when available. Diagnoses were extracted from pathology reports by the teledermatology service before transfer to study investigators. These diagnoses were then mapped to skin conditions by US board-certified dermatologists (including K.K. and S.J.H.). The case distribution across these diagnoses (both clinical and histologic) are presented in the Table; of 152 cases with available biopsy results, the diagnosis of 141 cases was growths.

## Study End Points

Our study was designed to evaluate 2 prespecified primary end points: (1) the agreement rate of the primary differential diagnosis of the PCPs with the reference diagnosis and (2) the agreement rate of the primary differential diagnosis of the NPs with the reference diagnosis. Based on the relative frequencies of conditions in this data set, the chance agreement is 3.77%.

Several secondary analyses were planned. First, for cases with biopsy results, diagnoses were classified as malignant, precancerous, or benign and were evaluated against biopsy-determined diagnoses. Clinicians were also asked to report whether they would have recommended a biopsy or referred the case to a dermatologist. For the subset of reads in which clinicians reported they would not opt for a referral, we assessed the diagnostic agreement rate. We also analyzed the time taken to review cases and self-reported diagnostic confidence.

Finally, 2 additional metrics (top-3 agreement and average overlap)[31] were used for more comprehensive evaluation of cases in which additional follow-up may be needed to arrive at a definitive diagnosis (Additional Evaluation Metrics section in the eMethods in the Supplement). An exploratory analysis also measured the effect of AI assistance on dermatologist agreement with reference diagnoses.

## Statistical Analysis

Data were analyzed from May 26, 2020, to January 27, 2021. To compare clinicians reviewing cases with AI assistance and reviewing cases without, we used a permutation test[32] with 1000 iterations. In each iteration, we permuted the assignment of whether reads were assisted or unassisted (ie, one-half of the full set of assisted and unassisted reads per case were selected to be assisted and the other half unassisted). Sensitivity analysis using a permutation test that preserved the reader cohorts' structure and another statistical analysis via a generalized linear mixed model produced similar results (see the Alternative Statistical Analyses section in the eMethods in the Supplement). Because this study had 2 prespecified primary end points (both 1-tailed superiority tests), we applied the Bonferroni correction, and *P* < .0125 was considered statistically significant (halved from α = 0.05 owing to 1-tailed tests and halved again owing to having 2 primary end points). Confidence intervals were computed by bootstrapping across both cases and readers for each sampled case (1000 iterations; sampling both cases and reader with replacement in each iteration). Hypothesis tests were conducted in Python, version 3.6.7 (Python Software Foundation).

## Results

This study involved the participation of 40 board-certified clinicians, including 20 PCPs (14 women [70.0%] and 6 men [30.0%]; mean experience, 11.3 [range, 2-32] years) who were located across 12 states and 20 NPs (18 women [90.0%] and 2 men [10.0%]; mean experience, 13.1 [range, 2-34] years) who practiced in primary care without physician supervision across 9 states. These clinicians reviewed 1048 teledermatology cases (672 women [64.2%] and 375 men [35.8%], with 1 missing; median age, 43 [interquartile range, 30-56] years) from 11 sites (Table) and provided 0 to 5 differential diagnoses per case (mean [SD], 1.6 [0.7]), for a total of 41 920 case reviews. Every PCP and NP reviewed each case only once, either with or without AI assistance (Figure 1).

Artificial intelligence assistance was associated with significantly higher top-1 agreement with the reference diagnosis (**Figure 2**A and eTable 3 in the Supplement). For PCPs, the increase in diagnostic agreement was 10% (95% CI, 8%-11%; *P* < .001), from 48% to 58%; for NPs, the improvement was 12% (95% CI, 10%-14%; *P* < .001), from 46% to 58%. Assistance was associated with improvements for all 40 readers, although the magnitude varied by reader (range, 2%-22%; median, 10%) (Figure 2B). Similar improvements were observed beyond the primary diagnosis based on the top-3 agreement, average overlap, per-condition sensitivity, and κ value (eFigures 4 and 5 and eTable 3 in the Supplement). In an exploratory analysis, 2 dermatologists' agreement with the reference diagnosis remained largely unchanged with AI assistance, increasing by 2% (95% CI, −1% to 5%), from 63% to 66% (eFigures 4 and 5 in the Supplement).[26]

For cases with available biopsy diagnoses (n = 141), the readers' accuracy at classifying lesions as malignant, precancerous, or benign trended upward by 3% for PCPs (95% CI, −1% to 7%) from 64% to 67% and by 8% for NPs (95% CI, 3%-13%) from 60% to 68% (Figure 2C-D). Subgroup analysis further found that sensitivity for malignant lesions, precancerous lesions, infectious skin diseases, and categories of hair loss trended upward or remained similar with assistance for both NPs and PCPs, with improvements ranging from −1% to 36% (eTable 4 in the Supplement).

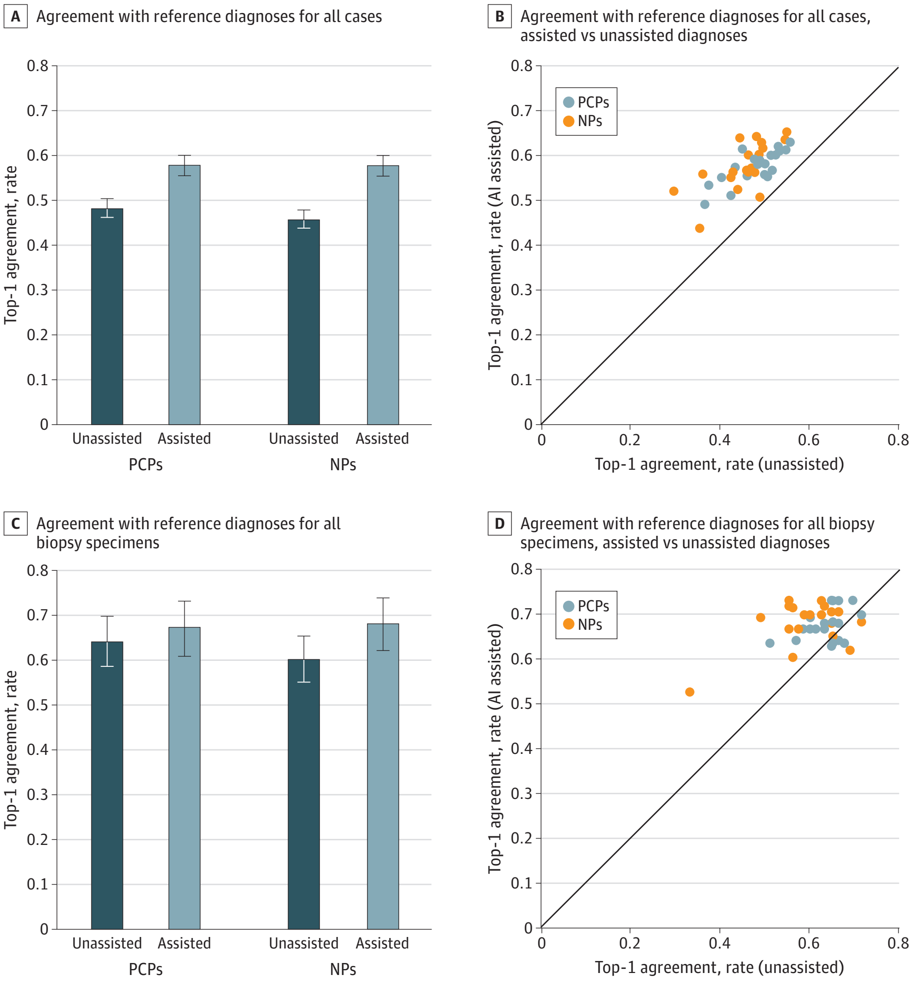
On the subset of cases in which the top prediction of AI was accurate (63% of cases), the use of assistance was associated with an increased top-1 agreement with reference diagnosis of 18% (95% CI, 16%-20%) for PCPs and 21% (95% CI, 19%-23%) for NPs. On the contrary, when none of the AI tool's predictions was correct (13% of cases), the agreement was 8% lower (95% CI, 5%-12%) for PCPs and 9% lower (95% CI, 6%-12%) for NPs. The effects were intermediate when the correct diagnosis was in the second or third position instead of the first (see the Impact of AI Accuracy on Assistance section of eMethods and eFigures 6 and 7 in the Supplement). An exploratory analysis also suggested that assistance was particularly beneficial for less ambiguous cases. For example, in the subset of cases in which the dermatologist panel had unanimous agreement, the use of AI assistance was associated with a top-1 agreement increase of 13% (95% CI, 10%-15%) for PCPs and of 16% (95% CI, 14%-19%) for NPs (eFigure 8 in the Supplement). Subanalyses also indicated that assistance-associated benefits were consistent during the study and across several skin types (eFigures 9 and 10 in the Supplement).

Artificial intelligence assistance was also associated with changes in several simulated clinical decisions (**Figure 3**A-B). The rates of indicating a need for biopsy were 1% lower (95% CI, 0%-3%) for

PCPs and 2% lower (95% CI, 1%-3%) for NPs; the rate of desire was 3% lower (95% CI, 1%-4%) for both PCPs and NPs (eTable 5 in the Supplement). For cases in which readers indicated referrals were unnecessary, their top-1 agreement rate with dermatologists was higher by 10% for PCPs (95% CI, 8%-12%), from 51% to 61%, and by 12% for NPs (95% CI, 10%-14%), from 47% to 59%, with a similar effect on referred cases (Figure 3C-D and eFigure 11 in the Supplement). In related findings, self-reported diagnostic confidence was substantially higher with AI assistance for both reader cohorts (**Figure 4**A). The top-1 agreement of cases that were rated with more than 90% confidence was substantially higher (73% vs 64% for PCPs and 68% vs 58% for NPs) (eFigure 12 in the Supplement).

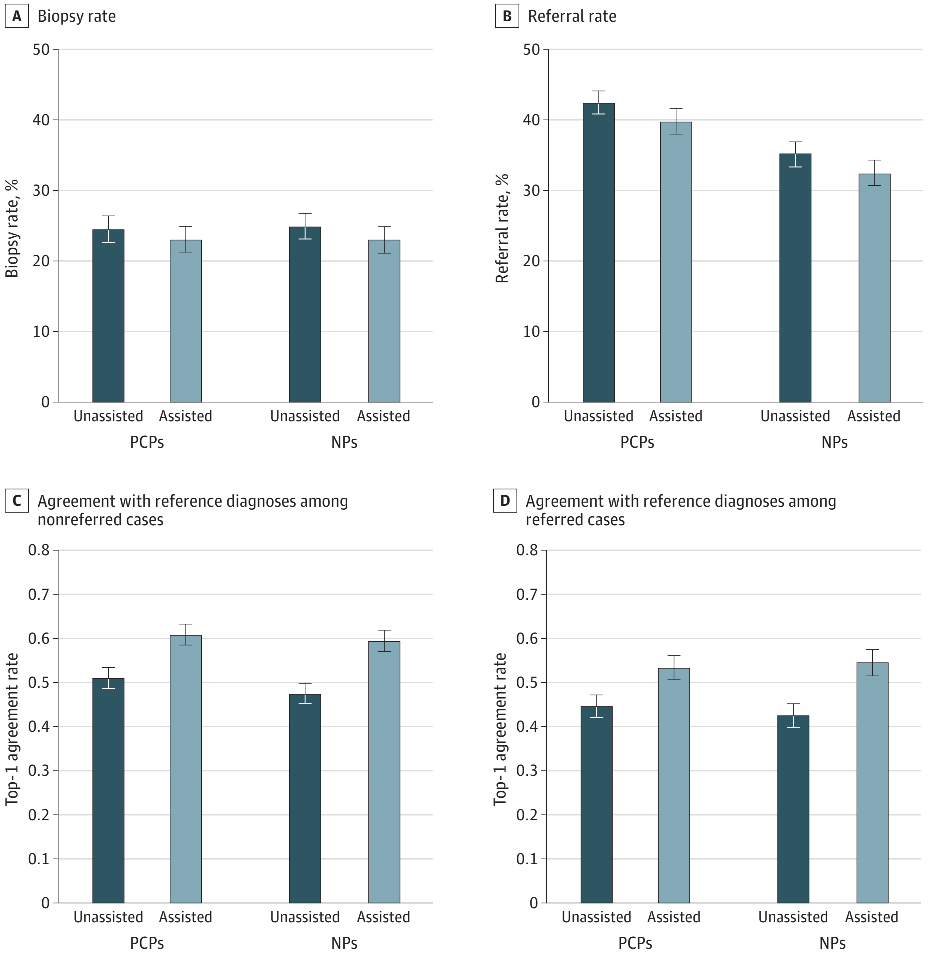
In terms of review time per case, AI assistance was associated with a slightly increased median review time. A difference of 5 (95% CI, 0-8) seconds, from 89 to 94 seconds, was observed for PCPs and a difference of 7 (95% CI, 5-10) seconds, from 77 to 84 seconds, was observed for NPs (Figure 4B and eFigure 9D in the Supplement). We also present representative examples of cases in which AI assistance was associated with the largest increases or decreases in agreement with reference diagnoses (eFigures 13 and 14 in the Supplement) and results of follow-up surveys investigating the usefulness of various AI assistant features (eFigures 15-17 in the Supplement).

Figure 2. Comparison of Clinicians' Diagnostic Agreement Rate With Dermatologists When Assisted by Artificial Intelligence (AI) vs Unassisted



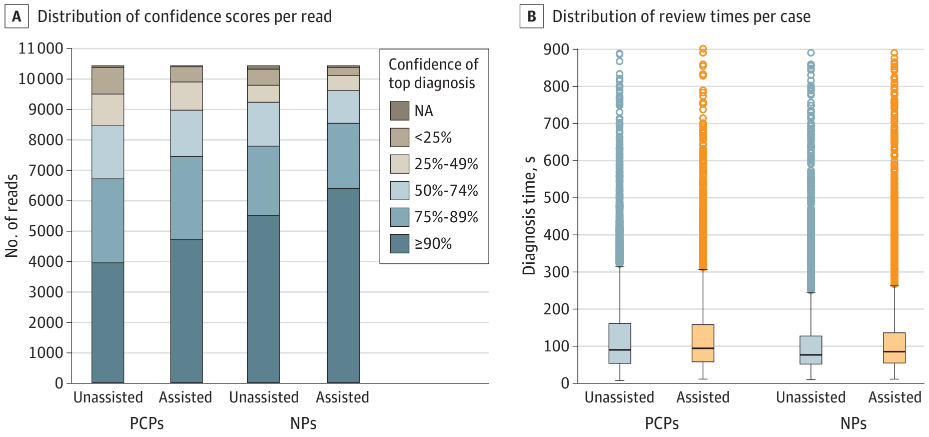
Every clinician (primary care physicians [PCPs] or nurse practitioners [NPs]) provided their differential diagnosis (several rank-ordered conditions), which were then mapped to 419 skin conditions. Only agreement in the top differential diagnosis (how often the clinicians' primary diagnosis agreed with the top diagnosis of a panel of dermatologists [top-1 agreement]) is considered, with additional details in eFigures 4 and 5 in the Supplement. Panels A and B cover all 1048 cases, whereas panels C and D cover 141 cases with growths and biopsy confirmation. A, Top-1 agreement increased with AI assistance (*P* < .001 for both PCPs and NPs). B, For top-1 agreement for unassisted vs assisted modalities for each individual clinician, a value above the diagonal indicates that the clinician had a higher agreement with dermatologists when assisted by AI. C and D, A similar analysis evaluated diagnostic accuracy for growths with biopsy confirmation on the 3-way classification of malignant, precancerous, and benign. Error bars represent 95% CIs. Additional analysis of assistance stratified by AI agreement with the reference diagnoses is presented in eFigures 6 and 7 in the Supplement.

Figure 3. Comparing Simulated Clinical Decisions by Clinicians When Assisted by Artificial Intelligence vs Unassisted



A, Rate of biopsy for all cases. B, Rate of referrals for all cases. C, Diagnostic accuracy among nonreferred cases. D, Diagnostic accuracy among referred cases. Top-3 agreement rates for cases for whom the primary care physicians (PCPs) and nurse practitioners (NPs) did and did not indicate a referral are presented in eFigure 11 in the Supplement. Error bars represent 95% CIs.

Figure 4. Comparing Clinicians' Confidence and Case Review Time When Assisted by Artificial Intelligence vs Unassisted



A, Confidence of the primary care physicians (PCPs) and nurse practitioners (NPs) as a stacked bar plot. NA indicates cases for which the clinician could not provide a diagnosis. B, Comparison of the differences in case review time for the full set of 1048 cases as a box plot. The box edges represent quartiles, whereas the whiskers extend to the last observed points that fall within 1.5 times the interquartile range from the quartiles. Outliers beyond the whiskers are indicated with dots; a total of 182 (0.4% of the reads) outliers beyond 900 seconds are excluded from the 4 box plots for ease of visualization. The median time for diagnosis increased from 89 to 94 seconds for PCPs and from 77 to 84 seconds for NPs.

## Discussion

In this study, 40 clinicians each reviewed 1048 teledermatology cases, with AI assistance for a random half of the cases and without AI assistance for the remaining half. Artificial intelligence assistance was associated with a higher agreement rate with dermatologists' reference diagnoses for both PCPs and NPs. The absolute effect size of 10% and 12% corresponds to an improved diagnosis for 1 in every 8 to 10 cases.

For both PCPs and NPs, AI assistance was also associated with lower rates of recommending a biopsy or specialist referral, marked increase in self-reported diagnostic confidence, and higher diagnostic agreement rates (with dermatologists) in nonreferred cases. These observations suggest that AI assistance improved skin condition diagnosis and diagnostic confidence of nonspecialists without incurring a reflexive increased use of referrals or biopsies. These improvements came at a modest cost of only a median of 5 to 7 additional seconds per case.

Our observations suggest that AI has the potential to augment the ability of PCPs and NPs independently practicing primary care to diagnose and triage skin conditions more effectively. Cutaneous disease is the chief complaint in 12% to 21% of primary care visits,[33-36] and access to dermatologists is limited. Nonspecialists have suboptimal diagnostic accuracy and have been shown to perform more biopsies while diagnosing fewer malignant neoplasms than dermatologists.[37] Therefore, improving the diagnostic accuracy of nonreferred cases while reducing unnecessary referrals and biopsies could have enormous implications for health care systems.

According to the American Academy of Dermatology,[38] the estimated direct health care cost of skin disease in the US is $75 billion, including $46 billion in medical costs (office visits, procedures, and tests), with an additional $11 billion of indirect opportunity costs from missed work or decreased productivity for patients and their caregivers. Appropriate diagnosis of dermatologic conditions at the point of care in primary care settings could translate to fewer delays in diagnosis and management and increased capacity for dermatology offices. Artificial intelligence also has the potential to enhance triage by improving the quality of information in referrals and enable dermatology offices to better prioritize the urgency of referrals. The clinical impact of this tool would need to be determined in prospective studies.

This AI tool uses as input images of the skin condition as well as a structured medical history. These images were taken using consumer-grade point-and-shoot cameras and mobile devices without specialized hardware. The interface used in this study was designed for store-and-forward teledermatology; however, extension to live, interactive teledermatology is in principle straightforward. In either case, the telemedicine format could be particularly useful in the COVID-19 era[39] for populations at high risk of complications in the event of infection due to in-person care. The AI tool could also be used in an in-person clinic setting because AI interpretation of images is feasible within seconds on modern smartphones. Such use could enable physicians to conduct follow-up tests (eg, potassium hydroxide test to confirm fundal infection), ask clarifying questions about the medical history, or conduct a closer physical examination to realize greater improvements in diagnostic ability.

More generally, and consistent with the consensus statements from both the American Medical Association[40] and the American Academy of Dermatology,[41] this tool was specifically designed to augment clinicians' diagnostic ability. To improve trust and empower readers to evaluate suggestion reliability, the tool provides a measure of its confidence and canonical examples of each suggested diagnosis. For skin conditions from which the AI algorithm had limited data to learn, suggestions are accompanied by a limited data warning. These features were designed to enable nonspecialists to diagnose cases more accurately and with greater confidence.

Other studies have explored the potential of AI-based dermatology tools. Han et al[15] found a 7% increase in diagnostic accuracy when 2 dermatologists and 2 residents reviewed 2201 cases a second time with AI assistance. Assistance-associated improvements were also seen for 21 dermatologists and 26 residents on 240 images for detection of malignant neoplasms.[15] Tschandl et al[24] highlighted the importance of effective human/computer interaction for AI tools for interpreting dermoscopic images, with improvements in showing multiclass prediction probabilities by skin condition but not for binary predictions of malignant neoplasms or AI-based retrieval of similar images. Our study complements these prior works. First, we evaluated images from nonspecialized, widely available devices. Second, we specifically examined the effect of AI assistance on PCPs and NPs, who perform most skin condition assessments. In addition, we assessed 2 pivotal clinical decisions: biopsy and referral. Finally, our randomized study design avoids any potential memory effects of reviewing the same case more than once.

### Limitations

This study has some limitations. First, these were teledermatology cases that were a mix of cases that were referred from primary care and other cases that were submitted at the patient's request. The potentially increased case difficulty and case enrichment may have affected clinician diagnostic performance. Second, in terms of Fitzpatrick skin types[42] (which categorize skin tone and propensity to tan), types I and V are underrepresented, and type VI is absent in this data set.[26] Because disease can present differently across skin types, the further study of additional skin types is warranted. Third, AI-associated improvements for malignant neoplasms were lower than those across all cases, and future work is needed to further improve the AI tool for malignant neoplasms. Our randomized study design of 1 modality per case/reader pair precludes inferences about any specific case and reader. Alternative study designs such as sequential reading (unassisted followed by assisted) or fully crossed setups could be explored, although biases from anticipation of AI assistance or incomplete washout will need to be averted.[28] Finally, the "store-and-forward" nature of these cases restricted the ability of the clinicians to ask follow-up questions and perform tests. As such, the insights here are more directly relevant to a store-and-forward setting than in-person clinics or live interactive telemedicine visits.

## Conclusions

Our AI tool was significantly associated with improved PCP and NP diagnostic agreement with dermatologists on skin condition cases from a teledermatology service. Prospective studies are warranted to study the impact of its use in both telemedicine settings and in-person primary care visits.

### ARTICLE INFORMATION

**Accepted for Publication:** February 1, 2021.

**Published:** April 28, 2021. doi:10.1001/jamanetworkopen.2021.7249

**Corresponding Author:** Yun Liu, PhD, Google Health, 3400 Hillview Ave, Palo Alto, CA 94304 (liuyun@google.com).

**Author Affiliations:** Google Health, Palo Alto, California (Jain, Way, Gupta, Gao, de Oliveira Marinho, Hartford, Sayres, Eng, Nagpal, DeSalvo, Corrado, Peng, Webster, Dunn, Coz, Yun Liu, Bui, Yuan Liu); Google Health via Advanced Clinical, Deerfield, Illinois (Kanada, Huang); Division of Hospital Medicine, University of California, San Francisco (Bui).

**Author Contributions:** Drs Bui and Yuan Liu contributed equally to the study. Mr Jain and Dr Yun Liu had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Jain, Way, de Oliveira Marinho, Sayres, Eng, Corrado, Peng, Webster, Dunn, Coz, Huang, Yun Liu, Bui, Yuan Liu.

*Acquisition, analysis, or interpretation of data:* Jain, Way, Gupta, Gao, Hartford, Sayres, Kanada, Nagpal, DeSalvo, Dunn, Coz, Huang, Yun Liu, Bui, Yuan Liu.

*Drafting of the manuscript:* Jain, Gupta, de Oliveira Marinho, Hartford, Sayres, Nagpal, Yun Liu, Bui, Yuan Liu.

*Critical revision of the manuscript for important intellectual content:* Jain, Way, Gao, Sayres, Kanada, Eng, Nagpal, DeSalvo, Corrado, Peng, Webster, Dunn, Coz, Huang, Yun Liu, Bui, Yuan Liu.

*Statistical analysis:* Jain, Way, Gupta, Gao, Yun Liu, Yuan Liu.

*Obtained funding:* Corrado, Peng, Webster, Dunn, Bui, Yuan Liu.

*Administrative, technical, or material support:* Way, Gupta, de Oliveira Marinho, Hartford, Kanada, Eng, DeSalvo, Peng, Dunn, Coz, Huang, Yun Liu, Bui, Yuan Liu.

*Supervision:* Corrado, Peng, Webster, Coz, Huang, Yun Liu, Bui, Yuan Liu.

**Conflict of Interest Disclosures:** Mr Jain reported a patent pending and ownership of Alphabet stock. Mr Way reported a patent pending and ownership of Alphabet stock. Mrs Gupta reported a patent pending and ownership of Alphabet stock. Dr Gao reported ownership of Alphabet stock. Mr de Oliveira Marinho reported ownership of Alphabet stock. Mr Hartford reported ownership of Alphabet stock. Dr Sayres reported a patent pending and ownership of Alphabet stock. Dr Kanada reported paid consulting for Google during the conduct of the study. Dr Eng reported a patent pending and ownership of Alphabet stock. Mr Nagpal reported ownership of Alphabet stock. Dr DeSalvo reported ownership of Alphabet stock. Dr Corrado reported ownership of Alphabet stock. Dr Peng reported ownership of Alphabet stock. Dr Webster reported ownership of Alphabet stock. Mr Dunn reported a patent pending and ownership of Alphabet stock. Mr Coz reported a patent pending and ownership of Alphabet stock. Dr Huang reported paid consulting for Google during the conduct of the study. Dr Yun Liu reported multiple patents pending and ownership of Alphabet stock. Dr Bui reported a patent pending and ownership of Alphabet stock. Dr Yuan Liu reported a patent pending and ownership of Alphabet stock. No other disclosures were reported.

**Funding/Support:** This study was supported by Google LLC.

**Role of the Funder/Sponsor:** Google LLC was involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** Sara Gabriele, BS, David Yen, BA, T. Saensuksopa, MHCI, ME, Carrie J. Cai, PhD, William Chen, BA, Quang Duong, PhD, Miles Hutson, BS, Dennis Ai, MS, Aaron Loh, MS, Bilson Campana, PhD, Jonathan Deaton, MS, Vivek Natarajan, MS, Ignacio Blanco, BS, Christopher Semturs, MS, Jessica Gallegos, MBA, Anita Misra, BTech, Roy Lee, BS, and all employees of Google LLC provided technical advice, discussion, and support. Peter Schalock, MD, and Sabina Bis, MD, both consultants for Google Health via Advanced Clinical, assisted with mapping free-text diagnoses to the structured list of 419 conditions. Justin Ko, MD, MBA, and Steven Lin, MD, Stanford Health Care, provided helpful discussions. No one was financially compensated for the stated contribution aside from their standard salary and associated compensation.

## REFERENCES

**1**. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7

**2**. Chen SC, Pennie ML, Kolm P, et al. Diagnosing and managing cutaneous pigmented lesions: primary care physicians versus dermatologists. *J Gen Intern Med*. 2006;21(7):678-682. doi:10.1111/j.1525-1497.2006.00462.x

**3**. Goulding JMR, Levine S, Blizard RA, Deroide F, Swale VJ. Dermatological surgery: a comparison of activity and outcomes in primary and secondary care. *Br J Dermatol*. 2009;161(1):110-114. doi:10.1111/j.1365-2133.2009.09228.x

**4**. Federman DG, Concato J, Kirsner RS. Comparison of dermatologic diagnoses by primary care practitioners and dermatologists: a review of the literature. *Arch Fam Med*. 1999;8(2):170-172. doi:10.1001/archfami.8.2.170

**5**. Pennie ML, Soon SL, Risser JB, Veledar E, Culler SD, Chen SC. Melanoma outcomes for Medicare patients: association of stage and survival with detection by a dermatologist vs a nondermatologist. *Arch Dermatol*. 2007;143(4):488-494. doi:10.1001/archderm.143.4.488

**6**. Feldman SR, Fleischer AB Jr, Williford PM, White R, Byington R. Increasing utilization of dermatologists by managed care: an analysis of the National Ambulatory Medical Care Survey, 1990-1994. *J Am Acad Dermatol*. 1997;37(5, pt 1):784-788. doi:10.1016/S0190-9622(97)70118-X

**7**. Viola KV, Tolpinrud WL, Gross CP, Kirsner RS, Imaeda S, Federman DG. Outcomes of referral to dermatology for suspicious lesions: implications for teledermatology. *Arch Dermatol*. 2011;147(5):556-560. doi:10.1001/archdermatol.2011.108

**8**. Moreno G, Tran H, Chia ALK, Lim A, Shumack S. Prospective study to assess general practitioners' dermatological diagnostic skills in a referral setting. *Australas J Dermatol*. 2007;48(2):77-82. doi:10.1111/j.1440-0960.2007.00340.x

**9**. Tran H, Chen K, Lim AC, Jabbour J, Shumack S. Assessing diagnostic skill in dermatology: a comparison between general practitioners and dermatologists. *Australas J Dermatol*. 2005;46(4):230-234. doi:10.1111/j.1440-0960.2005.00189.x

**10**. Federman DG, Kirsner RS. The abilities of primary care physicians in dermatology: implications for quality of care. *Am J Manag Care*. 1997;3(10):1487-1492.

**11**. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. doi:10.1038/nature21056

**12**. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One*. 2018;13(1):e0191493. doi:10.1371/journal.pone.0191493

**13**. Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images. In: Liebe B, Matas J, Seba N, Welling M, eds. *Computer Vision–ECCV 2016*. Springer; 2016:206-222. *Lecture Notes in Computer Science*; vol 9910. https://link.springer.com/chapter/10.1007/978-3-319-46466-4_13

**14**. Derm101.com. Derm101. Accessed August 9, 2019. https://www.derm101.com/

**15**. Han SS, Park I, Eun Chang S, et al. Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. *J Invest Dermatol*. 2020;140(9):1753-1761. doi:10.1016/j.jid.2020.01.019

**16**. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. *Med Image Comput Comput Assist Interv*. 2013;16(pt 2):403-410. doi:10.1007/978-3-642-40763-5_50

**17**. Codella NCF, Gutman D, Emre Celebi M, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). In: 2018 *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE; 2018. https://faculty.uca.edu/ecelebi/documents/ISBI_2018.pdf

**18**. Yuan Y, Chao M, Lo YC. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans Med Imaging*. 2017;36(9):1876-1886. doi:10.1109/TMI.2017.2695227

**19**. Haenssle HA, Fink C, Schneiderbauer R, et al; Reader Study Level-I and Level-II Groups. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166

**20**. Brinker TJ, Hekler A, Enk AH, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Cancer*. 2019;113:47-54. doi:10.1016/j.ejca.2019.04.001

**21**. Maron RC, Weichenthal M, Utikal JS, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer*. 2019;119:57-65. doi:10.1016/j.ejca.2019.06.013

**22**. Okuboyejo DA, Olugbara OO, Odunaike SA. Automating skin disease diagnosis using image classification. In: Ao SI, Douglas C, Grundfest WS, Brugstone J, eds. *Proceedings of the World Congress on Engineering and Computer Science*. Vol 2. Newstand Limited; 2013:850-854. http://www.iaeng.org/publication/WCECS2013/

**23**. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*. 2019;20(7):938-947. doi:10.1016/S1470-2045(19)30333-X

**24**. Tschandl P, Rinner C, Apalla Z, et al. Human-computer collaboration for skin cancer recognition. *Nat Med*. 2020;26(8):1229-1234. doi:10.1038/s41591-020-0942-0

**25**. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi:10.1136/bmj.h5527

**26**. Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med*. 2020;26(6):900-908. doi:10.1038/s41591-020-0842-3

**27**. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. In: Lampinen A, Gergle D, Shamma DA. *Proceedings of the ACM on Human-Computer Interaction*. Association for Computing Machinery; 2019;3(CSCW):1-24. doi:10.1145/3359206

**28**. Gallas BD, Chan HP, D'Orsi CJ, et al. Evaluating imaging and computer-aided detection and diagnosis devices at the FDA. *Acad Radiol*. 2012;19(4):463-477. doi:10.1016/j.acra.2011.12.016

**29**. Eadie LH, Taylor P, Gibson AP. Recommendations for research design and reporting in computer-assisted diagnosis to facilitate meta-analysis. *J Biomed Inform*. 2012;45(2):390-397. doi:10.1016/j.jbi.2011.07.009

**30**. Barnett ML, Boddupalli D, Nundy S, Bates DW. Comparative accuracy of diagnosis by collective intelligence of multiple physicians vs individual physicians. *JAMA Netw Open*. 2019;2(3):e190096. doi:10.1001/jamanetworkopen.2019.0096

**31**. Eng C, Liu Y, Bhatnagar R. Measuring clinician-machine agreement in differential diagnoses for dermatology. *Br J Dermatol*. 2020;182(5):1277-1278. doi:10.1111/bjd.18609

**32**. Droge B. Phillip Good: permutation, parametric, and bootstrap tests of hypotheses. *Metrika*. 2006;64(2):249-250. doi:10.1007/s00184-006-0088-1

**33**. Lowell BA, Froelich CW, Federman DG, Kirsner RS. Dermatology in primary care: prevalence and patient disposition. *J Am Acad Dermatol*. 2001;45(2):250-255. doi:10.1067/mjd.2001.114598

**34**. Verhoeven EWM, Kraaimaat FW, van Weel C, et al. Skin diseases in family medicine: prevalence and health care use. *Ann Fam Med*. 2008;6(4):349-354. doi:10.1370/afm.861

**35**. Sari F, Brian B, Brian M. Skin disease in a primary care practice. *Skinmed*. 2005;4(6):350-353. doi:10.1111/j.1540-9740.2005.04267.x

**36**. Britt H, Miller GC, Henderson J, et al. *General Practice Activity in Australia 2015-16: BEACH: Bettering the Evaluation and Care of Health*. Family Medicine Research Centre; 2016.

**37**. Anderson AM, Matsumoto M, Saul MI, Secrest AM, Ferris LK. Accuracy of skin cancer diagnosis by physician assistants compared with dermatologists in a large health care system. *JAMA Dermatol*. 2018;154(5):569-573. doi:10.1001/jamadermatol.2018.0212

**38**. Lim HW, Collins SAB, Resneck JS Jr, et al. The burden of skin disease in the United States. *J Am Acad Dermatol*. 2017;76(5):958-972.e2. doi:10.1016/j.jaad.2016.12.043

**39**. Gupta R, Ibraheim MK, Doan HQ. Teledermatology in the wake of COVID-19: advantages and challenges to continued care in a time of disarray. *J Am Acad Dermatol*. 2020;83(1):168-169. doi:10.1016/j.jaad.2020.04.080

**40**. American Medical Association. Augmented intelligence in health care policy report. June 2018. Accessed June 3, 2020. https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf

**41**. American Academy of Dermatology. American Academy of Dermatology position statement on augmented intelligence (AuI). May 18, 2019. Accessed June 3, 2020. https://server.aad.org/Forms/Policies/Uploads/PS/PS-Augmented%20Intelligence.pdf

**42**. Sachdeva S. Fitzpatrick skin typing: applications in dermatology. *Indian J Dermatol Venereol Leprol*. 2009;75(1):93-96. doi:10.4103/0378-6323.45238

**SUPPLEMENT.**

**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, Erythematosquamous and Papulosquamous Skin Disease, and Hair Loss
**eTable 5.** Rates of Recommending a Biopsy and Referral for PCPs and NPs