International Journal of Medicine
pISSN 1468-3814 | eISSN 2667-7008
https://theinternationalmedicine.com/

Original Researcher Article *DOI: 10.61336/im/26-11-02*

Evaluation of Diagnostic Accuracy of AI-Based Dermatology Applications in Patients with Fitzpatrick Skin Types IV–VI: A Hospital-Based Cross-Sectional Observational Study.

[1*]Dr. Aishwarya Maradi, [2]Dr. Shraddha Prabha, [3]Dr. Yashwanth kumar D

[1]*Senior Resident, Department of Dermatology Venereology and Leprosy, Esic Medical College Kalaburagi.*

[2]*Consultant Dermatologist Private Practice.*

[3]*Consultant Dermatologist, Private Practice.*

Correspondence:
Dr. Aishwarya Maradi.
1Senior Resident, Department of Dermatology Venereology and Leprosy, Esic Medical College Kalaburagi.

Received: 18-09-2026
Revised: 03-10-2026
Accepted: 18-10-2026
Published: 02-11-2026

Abstract

Background: Artificial intelligence (AI) has emerged as a promising technology for the automated identification and classification of dermatological disorders. However, concerns remain regarding its diagnostic performance among individuals with darker skin phototypes, particularly Fitzpatrick skin types IV–VI, because of potential underrepresentation in training and validation datasets. Aim: To evaluate the diagnostic accuracy and agreement of AI-based dermatology applications in identifying common dermatological conditions among patients with Fitzpatrick skin types IV–VI. ***Materials and Methods***: This hospital-based cross-sectional observational study is designed to include 140 adult patients attending the dermatology outpatient department of a tertiary care teaching hospital over three months. Patients with acne vulgaris, dermatophytosis, psoriasis, eczema, or vitiligo will undergo standardized clinical photography. Images will be evaluated using selected AI-based dermatology applications, and the resulting diagnoses will be compared with independent dermatologist assessments and histopathological findings where clinically indicated. Diagnostic performance will be evaluated using sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy, and Cohen's kappa coefficient. ***Results***: Results will be reported after completion of recruitment, image assessment, and statistical analysis. Planned analyses include disease-specific diagnostic performance, differences across Fitzpatrick skin types IV–VI, and agreement between AI-generated and dermatologist diagnoses. ***Conclusion***: This study is designed to provide evidence regarding the diagnostic performance and limitations of AI-based dermatology applications among individuals with darker skin phototypes. The findings may inform the responsible clinical implementation and validation of AI-assisted dermatological assessment.

Keywords: Artificial intelligence; dermatology; diagnostic accuracy; Fitzpatrick skin types; skin of color; machine learning; diagnostic agreement.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Artificial intelligence (AI) has emerged as an important technological advancement in modern healthcare, with applications ranging from diagnostic imaging and clinical decision support to disease prediction and personalized treatment [1]. Dermatology is particularly suitable for AI-assisted diagnosis because many dermatological conditions are identified primarily through visual examination of skin lesions [2]. Machine learning and deep learning algorithms, especially convolutional neural networks, have demonstrated the ability to recognize complex morphological patterns in clinical and dermoscopic images. These developments have generated considerable interest in the integration of AI-based applications into dermatological practice [3]. AI-assisted dermatology applications offer several potential advantages, including rapid preliminary assessment, improved access to dermatological expertise, and support for clinical decision-making [4]. Such tools may be particularly useful in resource-limited settings where specialist dermatological services are not readily available [5]. However, their clinical reliability depends substantially on the quality, diversity, and representativeness of the datasets used during algorithm development and validation [6]. Performance observed in selected image datasets may not translate directly into routine clinical practice, where differences in lesion morphology, lighting conditions, image quality, and patient characteristics influence diagnostic accuracy [7].

One major concern regarding AI-assisted dermatology is the potential for unequal diagnostic performance across different skin phototypes. The Fitzpatrick skin classification system categorizes skin into six phototypes according to its response to ultraviolet radiation [8]. Skin types IV, V, and VI generally represent individuals with greater constitutive pigmentation and varying tendencies toward tanning and burning [9]. Although this classification was originally developed to

characterize sun-reactive skin, it is frequently used in dermatological research to describe pigmentation-related patient groups [10]. The clinical appearance of several inflammatory, infectious, and pigmentary disorders may differ in individuals with darker skin. Erythema may be less visually apparent, while post-inflammatory hyperpigmentation, hypopigmentation, and alterations in pigmentation patterns may be prominent. These differences can complicate photographic interpretation and may influence the performance of AI algorithms trained predominantly on lighter skin images [11].

Recent investigations have emphasized the need to evaluate AI-based dermatology applications using diverse clinical image collections. Shah and Agarwal examined the diagnostic accuracy of AI applications on a diverse skin image dataset, highlighting the importance of representative evaluation [12]. Zaar and colleagues investigated the diagnostic performance of an online AI application for skin disease identification, contributing to evidence regarding the capabilities and limitations of automated diagnostic platforms [13]. Lau and colleagues specifically examined the diagnostic performance of ChatGPT in dermatology across Fitzpatrick phototypes and skin-of-color populations, reinforcing the importance of evaluating emerging AI technologies across heterogeneous patient groups [14]. In India, common dermatological conditions such as acne vulgaris, dermatophytosis, psoriasis, eczema, and vitiligo contribute substantially to outpatient dermatological consultations. These disorders demonstrate variable clinical morphology and pigmentation-associated features, making them relevant conditions for evaluating AI-based diagnostic applications among individuals with Fitzpatrick skin types IV–VI [15].

Despite increasing interest in AI-assisted dermatology, independent clinical validation in darker skin phototypes remains necessary. Diagnostic evaluation should extend beyond overall accuracy to include sensitivity, specificity, predictive values, and agreement with qualified dermatologists. Such analyses may identify disease-specific limitations and potential differences in diagnostic performance across skin phototypes [16]. Therefore, the present study aims to evaluate the diagnostic accuracy of selected AI-based dermatology applications among adult patients with Fitzpatrick skin types IV–VI attending a tertiary care dermatology outpatient department. By comparing AI-generated diagnoses with dermatologist assessments and histopathological findings wherever clinically indicated, the study seeks to generate evidence regarding the potential clinical usefulness and limitations of AI-assisted dermatological diagnosis.

MATERIALS AND METHODS

The present study is designed as a hospital-based cross-sectional observational investigation to be conducted in the Department of Dermatology, ESIC Medical College & Hospital, Kalaburagi, Karnataka, India. The study will assess the diagnostic accuracy of selected AI-based dermatology applications among patients with Fitzpatrick skin types IV–VI attending the dermatology outpatient department.

Study Duration

The proposed duration of the study is three months, including participant recruitment, clinical assessment, photographic documentation, AI-based evaluation, data collection, and statistical analysis.

Study Population

The study population will comprise adult patients aged 18 years and above who attend the dermatology outpatient department with common dermatological conditions, including acne vulgaris, dermatophytosis, psoriasis, eczema, and vitiligo.

Sample Size and Sampling Technique

A total of 140 patients are proposed for inclusion based on the sample size specified in the approved study planning documents. Participants will be selected using a convenience sampling technique. The precision of diagnostic accuracy estimates will depend on the final distribution of patients across individual disease categories and skin phototypes.

Inclusion Criteria

1. Patients aged 18 years and above.
2. Patients with Fitzpatrick skin types IV, V, or VI.
3. Patients clinically diagnosed with acne vulgaris, dermatophytosis, psoriasis, eczema, or vitiligo.
4. Patients willing to provide written informed consent.
5. Patients with lesions suitable for standardized clinical photography.

Exclusion Criteria

1. Patients unwilling to participate or provide informed consent.
2. Patients with poor-quality or unclear clinical photographs.
3. Patients with mixed or uncertain dermatological diagnoses.

4. Patients receiving treatment that has substantially altered lesion morphology.
5. Patients with severe systemic illness or dermatological emergencies.
6. Patients with incomplete clinical information required for diagnostic assessment.

Data Collection Procedure

Eligible patients will be approached during routine outpatient consultations. Following informed consent, demographic and clinical information will be recorded using a structured study proforma. Relevant variables will include age, sex, Fitzpatrick skin type, presenting complaints, duration of symptoms, clinical diagnosis, and pertinent dermatological history. Standardized clinical photographs will be obtained using a smartphone camera under consistent lighting and distance conditions. Images will focus on the affected skin region and will be evaluated for adequate clarity, lesion visibility, and suitability for AI analysis. The photographs will subsequently be processed using selected AI-based dermatology applications. The name, version, and operating conditions of each application will be documented before analysis. AI-generated diagnostic predictions will be recorded and compared with diagnoses established independently by qualified dermatologists. Histopathological confirmation will be incorporated in selected cases wherever clinically indicated. For reproducibility, the final manuscript will document whether the AI system provides a single diagnosis or a ranked list of differential diagnoses, the number of predictions considered correct, and the reference-standard procedure used to resolve diagnostic disagreements.

Outcome Measures

The primary outcome will be overall diagnostic accuracy of AI-based applications relative to the reference diagnosis. Secondary outcomes will include disease-specific sensitivity and specificity, positive and negative predictive values, diagnostic agreement using Cohen's kappa, and variation in diagnostic performance across Fitzpatrick skin types IV–VI.

Statistical Analysis

Data will be entered into Microsoft Excel and analyzed using SPSS version 26. Continuous variables will be summarized using means and standard deviations or appropriate nonparametric measures. Categorical variables will be presented as frequencies and percentages. Sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy will be calculated from diagnostic classification tables. Disease-specific analyses will use a one-versus-rest approach, where appropriate. Cohen's kappa coefficient will be used to quantify agreement beyond chance between AI-generated and dermatologist diagnoses. Confidence intervals will be reported for major diagnostic accuracy estimates. A p-value below 0.05 will be considered statistically significant for prespecified inferential comparisons.

Ethical Considerations

Written informed consent will be obtained before participation and clinical photography. Patient confidentiality will be maintained through anonymized data collection and secure storage of clinical information and images. The final manuscript will report the institutional ethics committee approval number and approval date after confirmation. Clinical images will be used in accordance with the consent obtained and applicable institutional policies.

RESULTS

Table 1. Demographic and Clinical Characteristics of Study Participants

Parameter	Category	Frequency (n)	Percentage (%)
Age (years)	18–30	48	34.3
	31–45	44	31.4
	46–60	32	22.9
	Above 60	16	11.4
Sex	Male	76	54.3
	Female	64	45.7
Fitzpatrick type	IV	65	46.4
	V	50	35.7
	VI	25	17.9
Clinical diagnosis	Acne vulgaris	35	25.0
	Dermatophytosis	32	22.9
	Psoriasis	26	18.6
	Eczema	27	19.3
	Vitiligo	20	14.3

Illustrative interpretation: The simulated sample comprises 140 adults, with the largest age group being 18–30 years (34.3%). Fitzpatrick skin type IV accounts for 46.4% of participants, followed by type V (35.7%) and type VI (17.9%). Acne vulgaris is the most frequent diagnosis (25.0%).

Table 2. Comparison of AI-Generated Diagnoses with Dermatologist Diagnoses

Dermatological condition	Total cases	Correct AI diagnoses	Incorrect AI diagnoses	Sensitivity (%)
Acne vulgaris	35	31	4	88.6
Dermatophytosis	32	27	5	84.4
Psoriasis	26	21	5	80.8
Eczema	27	21	6	77.8
Vitiligo	20	19	1	95.0
Total	140	119	21	85.0

Illustrative interpretation: The simulated AI system correctly classifies 119 of 140 patients, giving an overall top-1 diagnostic accuracy of 85.0%. Sensitivity is highest for vitiligo (95.0%) and lowest for eczema (77.8%).

Table 3. Disease-Specific Diagnostic Performance

Condition	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Acne vulgaris	88.6	97.1	91.2	96.2	95.0
Dermatophytosis	84.4	96.3	87.1	95.4	93.6
Psoriasis	80.8	96.5	84.0	95.7	93.6
Eczema	77.8	93.8	75.0	94.6	90.7
Vitiligo	95.0	97.5	86.4	99.2	97.1

Illustrative interpretation: The simulated results show variation in diagnostic performance between conditions. Vitiligo demonstrates the highest sensitivity, while eczema shows the lowest sensitivity and positive predictive value. These figures are internally consistent with a hypothetical five-class confusion matrix; they are not observed measurements.

Table 4. AI Diagnostic Accuracy Across Fitzpatrick Skin Types

Fitzpatrick skin type	Total patients	Correct diagnoses	Incorrect diagnoses	Accuracy (%)
Type IV	65	58	7	89.2
Type V	50	42	8	84.0
Type VI	25	19	6	76.0
Total	140	119	21	85.0

Illustrative interpretation: Diagnostic accuracy in this simulated example decreases from 89.2% among patients with Fitzpatrick skin type IV to 76.0% among those with type VI. Such a pattern, if observed in actual data, would require further analysis of uncertainty, disease mix, and potential confounding before attributing the difference to algorithmic bias.

Table 5. Agreement Between AI-Based and Dermatologist Diagnoses

Comparison	Observed agreement (%)	Cohen's kappa
Overall, AI versus dermatologist diagnosis	85.0	0.81
Fitzpatrick type IV	89.2	—
Fitzpatrick type V	84.0	—
Fitzpatrick type VI	76.0	—

Illustrative interpretation: Overall Cohen's kappa is approximately 0.81 for the simulated diagnostic classifications, indicating substantial to near-perfect agreement by commonly used interpretive conventions. Separate phototype-specific kappa coefficients require complete diagnostic cross-tabulations for each phototype and cannot be derived from accuracy totals alone.

Table 6. Distribution of AI Diagnostic Errors

Reference diagnosis	Most frequent incorrect AI diagnosis	Number of errors	Error rate (%)
Acne vulgaris	Eczema	4	11.4
Dermatophytosis	Eczema	5	15.6
Psoriasis	Eczema	5	19.2
Eczema	Psoriasis	6	22.2
Vitiligo	Eczema	1	5.0

DISCUSSION

Artificial intelligence has considerable potential to transform dermatological diagnosis by enabling automated interpretation of clinical images and providing rapid diagnostic suggestions. Nevertheless, its clinical applicability depends on the accuracy, consistency, and generalizability of diagnostic predictions across diverse patient populations. The present study is designed to investigate these issues among patients with Fitzpatrick skin types IV–VI, a population for which dedicated diagnostic validation is particularly important.

Previous investigations have demonstrated growing interest in evaluating AI-based dermatology systems using clinically diverse image datasets. Shah and Agarwal examined the diagnostic accuracy of AI applications on a diverse collection of skin images, emphasizing the relevance of dataset diversity when evaluating automated diagnostic technologies [17]. Similarly, Zaar and colleagues evaluated an online AI application for skin disease diagnosis, providing evidence that automated diagnostic tools require careful assessment before their integration into routine clinical workflows [18].

A central concern in dermatological AI research is the possibility that model performance may vary according to patient pigmentation and clinical presentation. Fitzpatrick skin classification provides one method of grouping patients according to skin phototype, although it is not a direct measure of constitutive skin color or ethnicity [19]. Differences in pigmentation, erythema visibility, lesion contrast, and post-inflammatory changes may influence the photographic appearance of dermatological conditions. Consequently, algorithms developed using insufficiently diverse training images may perform inconsistently when applied to populations that differ from those represented during model development.

Lau and colleagues investigated the diagnostic performance of ChatGPT across Fitzpatrick phototypes and skin-of-color populations, drawing attention to the need for systematic evaluation of emerging AI technologies in dermatology [20]. The present study builds on this area of investigation by focusing on common dermatological conditions encountered in a tertiary care outpatient setting.

The inclusion of acne vulgaris, dermatophytosis, psoriasis, eczema, and vitiligo is clinically relevant because these disorders exhibit different morphological and pigmentary characteristics. Acne vulgaris commonly presents with comedones and inflammatory lesions, whereas dermatophytosis may produce annular or scaly plaques. Psoriasis and eczema can exhibit overlapping inflammatory manifestations, while vitiligo is characterized by depigmented lesions. Such differences may affect diagnostic recognition by automated systems [21].

An important feature of the proposed study is its use of dermatologist diagnoses as the principal reference standard, supplemented by histopathological confirmation when clinically indicated. This approach will permit evaluation of AI performance against routine specialist assessment while recognizing that histopathology is not necessary or appropriate for every included disorder.

The planned assessment of sensitivity, specificity, predictive values, and Cohen's kappa will provide a broader understanding of diagnostic performance than overall accuracy alone. A system may achieve satisfactory overall accuracy while demonstrating limited sensitivity for individual conditions or particular skin phototypes. Therefore, disease-specific and phototype-specific analyses are essential for interpreting clinical utility [22].

The study has several anticipated limitations. Convenience sampling at a single tertiary care center may reduce generalizability. The proposed sample of 140 patients may also provide limited precision for individual disease categories, particularly if their distribution is uneven. Variations in photographic quality, AI application versions, diagnostic confidence, and the selection of a reference standard may influence the observed results.

Furthermore, the exclusion of unclear images and diagnostically uncertain cases may lead to performance estimates that do not fully reflect real-world clinical conditions. Differences across skin phototypes may also be confounded by variations in disease prevalence and lesion characteristics.

The eventual findings should therefore be interpreted cautiously and should not be used to support autonomous AI diagnosis without appropriate clinical validation. Instead, the study may provide useful evidence for improving inclusive dermatological datasets, refining AI diagnostic algorithms, and guiding the responsible integration of AI-assisted tools into specialist care.

CONCLUSION

Final conclusions regarding accuracy, reliability, and clinical applicability will depend on the actual study findings. Until those findings are available, AI-based dermatology applications should be considered potential decision-support tools rather than replacements for qualified dermatological assessment.

REFERENCES

1. Shah AK, Agarwal M. Diagnostic accuracy of artificial intelligence applications on a diverse skin image set. Cureus. 2026;18(1). doi:10.7759/cureus.102354.
2. Zaar O, Larson A, Polesie S, Saleh K, Tarstedt M, et al. Evaluation of the diagnostic accuracy of an online artificial intelligence application for skin disease diagnosis. Acta Derm Venereol. 2020;100(16). doi:10.2340/00015555-3624.

3. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988;124(6):869-871. doi:10.1001/archderm.1988.01670060015008.
4. Lau CB, Kwa M, Shen L, Smith GP. Assessing the diagnostic performance of ChatGPT in dermatology across Fitzpatrick phototypes and skin of color. J Am Acad Dermatol. 2025;92(3):578-579. doi:10.1016/j.jaad.2024.10.038.
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-118. doi: 10.1038/nature21056.
6. Daneshjou R, Vodrahalli K, Novoa RA, Jenkins M, Liang W, Rotemberg V, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022;8(32):eabq6147. doi: 10.1126/sciadv.abq6147.
7. Groh M, Badri O, Daneshjou R, Koochek A, Harris C, Soenksen LR, et al. Deep learning-aided decision support for diagnosis of skin disease across skin tones. Nat Med. 2024;30:573-583. doi: 10.1038/s41591-023-02728-3.
8. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. Nat Med. 2020;26:1229-1234. doi: 10.1038/s41591-020-0942-0.
9. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. Nat Med. 2020;26:900-908. doi: 10.1038/s41591-020-0842-3.
10. Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. JAMA Dermatol. 2021;157(11):1362-1369. doi: 10.1001/jamadermatol.2021.3129.
11. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol. 2018;154(11):1247-1248. doi: 10.1001/jamadermatol.2018.2348.
12. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J Invest Dermatol. 2018;138(7):1529-1538. doi: 10.1016/j.jid.2018.01.028.
13. Rotemberg V, Kurtansky N, Betz-Stablein B, Caffery L, Chousakos E, Codella N, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Sci Data. 2021;8:34. doi: 10.1038/s41597-021-00815-z.
14. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data. 2018;5:180161. doi: 10.1038/sdata.2018.161.
15. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, 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.
16. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer. 2019;119:11-17. doi: 10.1016/j.ejca.2019.05.023.
17. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. 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.
18. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. Eur J Cancer. 2019;111:148-154. doi: 10.1016/j.ejca.2019.02.005.
19. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, 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.
20. Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. Eur J Cancer. 2021;156:202-216. doi: 10.1016/j.ejca.2021.06.049.
21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi: 10.1136/bmj.h5527.
22. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6:e012799. doi: 10.1136/bmjopen-2016-012799.