Artificial intelligence-enhanced dermoscopy for early skin cancer detection in private practice: insights from a retrospective cohort study

Journal of Investigative Dermatology (2025) ■, ■−■; doi:10.1016/j.jid.2025.12.006

TO THE EDITOR

Given the rising incidence of skin cancer (Defossez et al, 2021), combined with prolonged wait times for dermatologic consultations and a declining dermatologist-to-population ratio (Perrot et al, 2024; Yadav et al, 2016), artificial intelligence (AI) may help shorten diagnostic delays by supporting general practitioners in improving their diagnostic capabilities. Deep learning, particularly convolutional neural networks (CNNs), now matches or surpasses clinicians in classifying skin lesions, with high sensitivity and specificity (Salinas et al, 2024). These tools are increasingly integrated into the clinical workflow. However, despite promising accuracy in specialized hospital settings, their real-world performance remains underexplored (Jones et al, 2022).

This retrospective, population-based study was conducted in a private dermatology and esthetic surgery center. Adults presenting with any skin lesions were included. All lesions were systematically imaged by the general practitioner using a dermoscopy camera (Medicam 1000s, Fotofinder GmbH) coupled with a CE-cleared CNN tool (Moleanalyzer Pro) independent of his initial clinical examination. Each lesion had dermoscopic images analyzed by the AI, which generated a malignancy score (0−1) and classified lesions as low risk (0.05−0.19), intermediate risk (0.20−0.49), or high risk (0.50−0.95). A score ≥0.2 was selected as the cut off for high-risk classification (Supplementary Materials and Methods). Biopsy decisions were based on the combination of clinical impression and AI data. Histopathology was assessed by 2 blinded dermatopathologists. Premalignant lesions were classified as malignant. Clinical variables were analyzed according to the ABCDE (Asymmetry, Borders, Color, Diameter, Evolution) rule to explore concordance between algorithmic and human visual assessment (Supplementary Materials and Methods).

Among 403 lesions analyzed in a predominantly light skin phototype (Fitzpatrick II/III) cohort with a median age of 48 years (Supplementary Table S1), the AI tool classified 273 as high risk (≥0.2) and 130 as low risk (<0.2) (Supplementary Table S2), whereas histopathology identified 104 malignant and 299 benign lesions (Supplementary Table S3). This yielded high sensitivity (92.3%) but modest specificity (40.8%), with an overall good area under the curve of 0.814 (95% confidence interval = 0.765−0.863). Increasing the malignancy threshold to 0.5 improved specificity (55.9%) but reduced sensitivity (81.7%), whereas the optimal Youden index (0.49) was achieved at 0.9, yielding 83.6% specificity and 65.4% sensitivity (Figure 1). Eight false negatives were misclassified by the AI: 5 actinic keratosis, 2 superficial spreading melanomas, and 1 squamous cell carcinoma.

In this cohort, patient age >50 years, recent lesion evolution, and diameter >6 mm were significantly associated with histopathologic malignancy (OR_{age} = 0.078, OR_{evol} = 1.96, OR_{dia} = 1.86; all $P < .001$). Analysis of ABCDE parameters showed that AI classification at the 0.2 threshold was primarily influenced by morphological features. Color variegation had the strongest association with AI positivity (86% vs 57%; OR = 4.58, $P < .001$), followed by border irregularity (81% vs 58%; OR = 3.12, $P < .001$), asymmetry (80% vs 63%; OR = 2.30, $P = .002$), and recent evolution (75% vs 59%; OR = 2.05, $P < .001$). Lesion diameter >6 mm was not significantly correlated with AI output (OR = 1.35, $P = .17$) (Supplementary Table S4).

When stratified by ABCDE score, AI sensitivity remained consistently high (≥0.92) across all subgroups at the 0.2 threshold, but specificity declined from 51.2% in ABCDE = 0−1 lesions to 13% in ABCDE = 4−5, suggesting increasing overclassification in clinically abnormal cases. Raising the threshold to 0.5 improved specificity (up to 64% for ABCDE = 0−1) but reduced sensitivity (∼0.56−0.87) (Figure 2 and Supplementary Table S5). Despite these threshold adjustments, overall accuracy remained stable across all ABCDE subgroups, indicating that the AI model's discriminative capacity was independent of clinical suspicion level on the basis of ABCDE criteria.

This study underscores the potential of AI-assisted dermoscopy to improve early skin cancer detection in a private medical center. Recent prospective studies evaluating AI in dermatology and primary care settings (Brancaccio et al, 2024) provide important benchmarks for widespread tools using CNNs. Our study, despite its retrospective design, aligns with the review's call for additional real-world evidence contributing to fill this gap by evaluating AI performance in a private clinic.

Although raising the Moleanalyzer Pro malignancy cut off to 0.5 or 0.9 improved specificity, it substantially reduced sensitivity. Given the clinical priority of minimizing missed malignancies, the 0.2 threshold appears most appropriate, although it leads to more false positives and limits the tool's use

Abbreviations: AI, artificial intelligence; CNN, convolutional neural network

Accepted manuscript published online XXX; corrected proof published online XXX

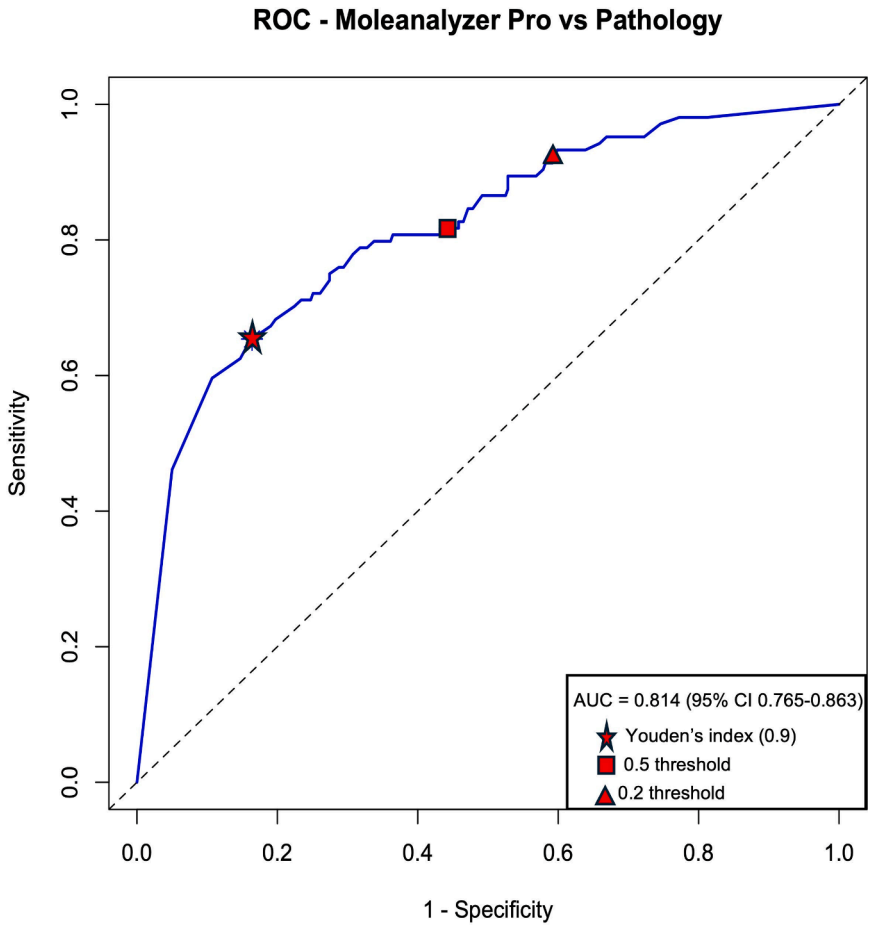

Figure 1. ROC curve of Moleanalyzer Pro diagnostic performance compared with pathology. ROC, receiver operator characteristic.

as a standalone diagnostic aid. In a recent meta-analysis, sensitivities and specificities using only clinical examination and images were 37.5% and 84.6%, respectively, for primary care physicians. When dermoscopy was added, diagnostic performance improved to 49.5% and 91.3%, respectively (Chen et al, 2025).

Accuracy remained consistent across all ABCDE levels, indicating that the CNN identifies subvisual cues not directly aligned with clinical heuristics. The weaker association of diameter and evolution with AI positivity does not necessarily represent a limitation but may reflect appropriate de-emphasis of less predictive, noncausal features. Prior explainable-AI studies similarly show that CNNs rely primarily on textural and pigment-gradient cues beyond the ABCDE rule (Giavina-Bianchi et al, 2023; Hauser et al, 2022).

Although our findings differ from those of experimental studies reporting better performances, the discrepancy likely reflects the transition from controlled image datasets to heterogeneous, real-life clinical material. Similar declines were reported in other real-world cohorts, yet the consistently high sensitivity supports AI-assisted triage as a safe tool for minimizing missed skin cancers (Miller et al, 2023). Moreover, biopsy decisions combined clinical judgment with the AI malignancy score. This reflects real-world use but represents a methodological limitation, potentially inflating sensitivity and reducing specificity.

Despite the single-center design and predominantly phototypes II–III population, our findings highlight the strong potential of AI-assisted dermoscopy as a decision-support tool for general practitioners with dermoscopy expertise. The CE-cleared CNN-based system demonstrated high sensitivity and consistent performance beyond traditional ABCDE criteria, capturing morphological cues often imperceptible to clinical assessment. Although newer transformer-based and hybrid models now outperform conventional CNNs (Cai et al, 2025), our results provide valuable real-world evidence on the clinical utility of a currently accessible AI tool. The general practitioner had a formal dermoscopy training in a dermatology-focused private medical centre. Although this may limit generalizability to standard primary care environments, such private-clinic populations remain largely underexplored in the dermatology-AI literature and, therefore, provide valuable real-world insight.

ETHICS STATEMENT

Ethical approval was obtained from the Paris Sorbonne University Research Ethics Committee (IRB00014283). Because this was a retrospective analysis of existing clinical images and deidentified data, the requirement for written, informed consent was waived by the institutional review board.

DATA AVAILABILITY STATEMENT

The dataset generated and analyzed during this study have been deposited in a publicly accessible online repository. The full anonymized dataset is available on the Mendeley Data repository: DOI 10.17632/pwbmn448wy.1

KEYWORDS

Artificial intelligence; Convolutional neural networks; Dermoscopy; Private practice; Skin cancer

ORCIDs

Raffaele Aguglia: http://orcid.org/0009-0007-1266-3409
Alexandre Lellouch: http://orcid.org/0000-0001-8191-8662

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

No funding was received for this study. We thank Manolo Boado and Sherwin Santiago for capturing all the lesions using the artificial intelligence tool. The study was conducted in a private dermatologic and esthetic surgery center in Paris, France.

AUTHOR CONTRIBUTIONS

Conceptualization: AGL, DMS, MA; Data Curation: RA, YD, LK, LB, LD, TC, GK; Formal Analysis: YD, RA, LVD; Investigation: RA, LVD, YB, HO; Methodology: AGL, MA, YD; Supervision: AGL, CLC, DMS; Validation: YD; Visualization: RA, LVD, YB; Writing – Original Draft Preparation: RA, MA, LVD, HO; Writing – Review and Editing: AGL, CC, DMS, YB

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at 10.1016/j.jid.2025.12.006.

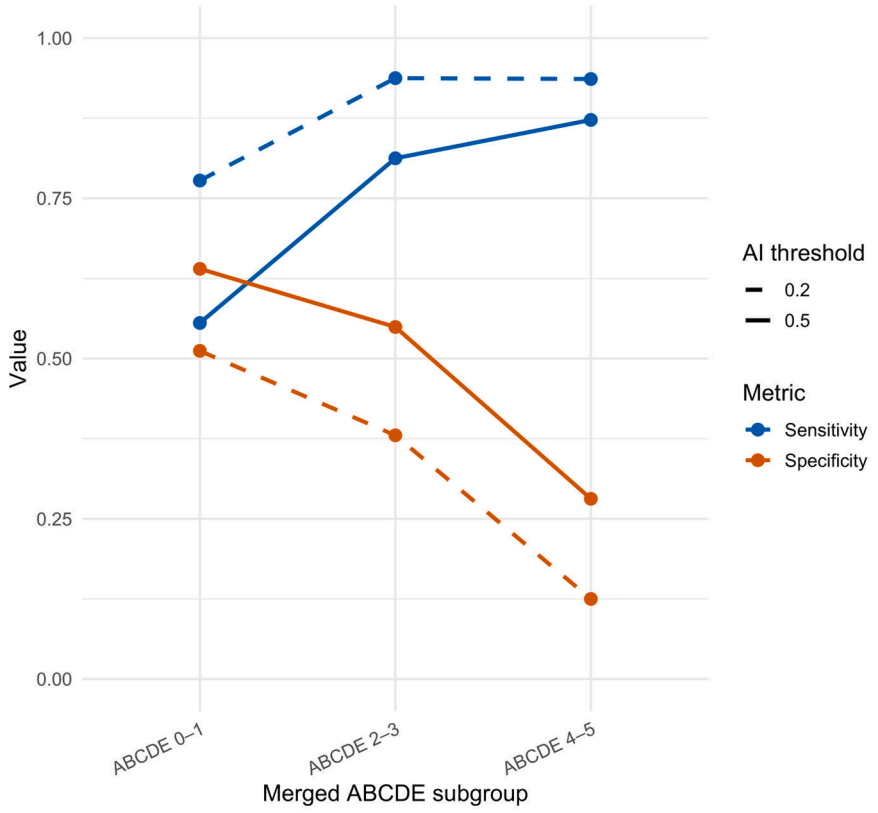

Figure 2. Evolution of sensitivity and specificity across ABCDE scenarios and 2 different malignancy thresholds. 0.2: dashed; 0.5: solid. ABCDE, Asymmetry, Borders, Color, Diameter, Evolution.

Raffaele Aguglia[1], Moshé Assouline[2], Loïc Van Dieren[3,4], Yanis Berkane[5,6], Haizam Oubari[7], Laura Bitton[8], Lucie Duverger[9], Theodoros Chrelias[2], Georgia Kanellopoulou[2], Leonard Knoedler[1], Curtis L. Cetrulo[1], David M. Smadja[10,11], Yohann Dabi[12] and Alexandre G. Lellouch[1,10,11,*]

[1]Department of Plastic, Reconstructive and Aesthetic Surgery, Cedars Sinai Hospital, Los Angeles, California, USA; [2]Maison Abeille Medical and Surgical Center for Dermatology & Aesthetics, Paris, France; [3]Vascularized Composite Allotransplantation Laboratory, Center for Transplantation Sciences, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; [4]Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium; [5]Department of Plastic, Reconstructive and Aesthetic Surgery, CHU Rennes, University of Rennes, Rennes, France; [6]SITI Laboratory, INSERM UMR 1236, Etablissement Francais du Sang, University of Rennes, Rennes, France; [7]Department of Plastic, Reconstructive and Aesthetic Surgery, Croix-Rousse Hospital, Hospices Civils de Lyon, Claude Bernard University Lyon 1, Lyon, France; [8]Medipath, Paris, France; [9]Centre National de Dermatopathologistes Indépendants, Ivry sur Seine, France; [10]INSERM Paris Cardiovascular Research Center, Paris City University, Paris, France; [11]Hematology Department, European Georges Pompidou Hospital, Assistance Publique Hôpitaux de Paris-Centre-Université de Paris Cité (AP-HP.CUP), Paris, France; and [12]Department of Obstetrics and Reproductive Medicine, Hôpital Tenon, Sorbonne University, Paris, France

[*]Corresponding author e-mail: alexandre.lellouch@cshs.org

REFERENCES

Brancaccio G, Balato A, Malvehy J, Puig S, Argenziano G, Kittler H. Artificial intelligence in skin cancer diagnosis: a reality check. J Invest Dermatol 2024;144:492–9.

Cai Y, Cai L, Song A, Yang C, Yang W, Yu R, et al. Transformer-assisted broad learning for hybrid intelligence-based skin cancer segmentation. Sci. Rep. Nature Publishing Group 2025;15:35532.

Chen JY, Fernandez K, Fadadu RP, Reddy R, Kim MO, Tan J, et al. Skin cancer diagnosis by lesion, physician, and examination type: a systematic review and meta-analysis. JAMA Dermatol 2025;161:135–46.

Defossez G, Uhry Z, Delafosse P, Dantony E, d'Almeida T, Plouvier S, et al. Cancer incidence and mortality trends in France over 1990–2018 for solid tumors: the sex gap is narrowing. BMC Cancer 2021;21:726.

Giavina-Bianchi M, Vitor WG, Fornasiero de Paiva V, Okita AL, Sousa RM, Machado B. Explainability agreement between dermatologists and five visual explanations techniques in deep neural networks for melanoma AI classification. Front Med (Lausanne) 2023;10:1241484.

Hauser K, Kurz A, Haggenmüller S, Maron RC, von Kalle C, Utikal JS, et al. Explainable artificial intelligence in skin cancer recognition: a systematic review. Eur J Cancer 2022;167:54–69.

Jones OT, Matin RN, van der Schaar M, Prathivadi Bhayankaram K, Ranmuthu CKI, Islam MS, et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. Lancet Digit Health 2022;4:e466–76.

Miller IJ, Stapelberg M, Rosic N, Hudson J, Coxon P, Furness J, et al. Implementation of artificial intelligence for the detection of cutaneous melanoma within a primary care setting: prevalence and types of skin cancer in outdoor enthusiasts. PeerJ 2023;11:e15737.

Perrot JL, Vélia K, Ripolles M, Kourdjee S, Freitas-Planello J, Venturi A, et al. Évolution de la prise en charge dermatologique dans un département français de 2013 à 2023. Ann Dermatol Venereol 2024;4:A275.

Salinas MP, Sepúlveda J, Hidalgo L, Peirano D, Morel M, Uribe P, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. NPJ Digit Med 2024;7:125.

Yadav G, Goldberg HR, Barense MD, Bell CM. A cross-sectional survey of population-wide wait times for patients seeking medical vs. cosmetic dermatologic care. PLoS One 2016;11:e0162767.

SUPPLEMENTARY MATERIALS AND METHODS

Study design

We designed a retrospective, population-based, real-life setting study conducted in a private dermatological and esthetic surgery center in Paris, France. Patients were included from November 1, 2022 to July 26, 2024. All patients received written information about the study, including their right to withdraw at any time. Ethical approval was obtained from the Paris Sorbonne University Research Ethics Committee (n° IRB00014283). This study was carried out according to the 2021 updated STARD-AI (Standards for Reporting of Diagnostic Accuracy Studies for Artificial Intelligence) and Strengthening the Reporting of Observational Studies in Epidemiology guidelines and the World Medical Association Declaration of Helsinki Ethical Principles (Sounderajah et al, 2021; von Elm et al, 2008; World Medical Association, 2013).

Participants and inclusion criteria

Records of nonopposing patients, aged >18 years, presenting with suspicious or nonsuspicious skin lesions, including suspected melanoma and non-melanoma skin cancers, were included. Both referred and nonreferred patients were eligible.

Imaging and artificial intelligence system

The diagnostic device used was an advanced dermoscopy camera (Medicam 1000s, Fotofinder Systems GmbH, Bad Birnbach, Germany) combined with a CE-cleared convolutional neural network system (Moleanalyzer Pro, Fotofinder Systems GmbH). The training and technical aspects of the artificial intelligence (AI) model has been previously described (Haenssle et al, 2018).

Patient workflow

The clinical examination is first conducted by the general practitioner who had a formal training in dermoscopy. Suspicious lesions are then captured either by the physician or the assistant and takes approximately 20 minutes. The images are then reviewed by the general practitioner. Each lesion had at least 1 macroscopic and 1 dermoscopic image captured (Medicam 1000s, ×20 magnification). After image capture, the Moleanalyzer Pro generated a malignancy risk score (0−1), categorizing lesions into low risk (0.05−0.19), intermediate risk (0.20−0.49), and high risk (0.50−0.95). The AI score is based solely on the dermoscopy image. The binary cut off of ≥0.2 corresponds to the manufacturer's recommended threshold for intermediate-risk classification. During analysis, alternative thresholds (0.5, corresponding to the "high-risk" threshold and 0.9 corresponding to the "optimal theoretical threshold") were evaluated to examine sensitivity−specificity trade offs, confirming that 0.2 offered the highest sensitivity and was thus the most clinically appropriate for a triage-oriented tool.

The decision to biopsy a lesion was made using both the general practitioner's clinical judgement and the AI malignancy score.

Histopathological confirmation

All biopsied lesions were analyzed by 2 independent, blinded pathologists (LVD or LB) specialized in skin cancer. Premalignant lesions, such as actinic keratoses, were classified as malignant for analysis purposes. In cases of initial disagreement, the slides were re-examined jointly in a consensus meeting. Final diagnoses were established through this consensus process, ensuring that no case was included without diagnostic agreement.

Data collection

Data were collected retrospectively and included patients' medical records; relevant medical history such as personal or familiar history of skin cancer; or any other risk factor related to skin malignancy, macroscopic and histological characteristics of the lesions, and AI scores. All included lesions had at least 1 clear digital and 1 dermoscopic photograph with AI scoring. Clinical features were assessed using the ABCDE (Asymmetry, Borders, Color, Diameter, Evolution) rule to standardize visual evaluation across all pigmented lesions. Although originally developed for melanoma detection, this framework was applied because over 80% of lesions in the cohort (334 of 403) were pigmented, providing a consistent and reproducible method for morphological classification.

Statistical analysis

Skin lesions classified as intermediate or high risk by the AI tool were positively considered "AI suspicious." Blinded histopathological reports were used as the gold standard for comparison. Data were analyzed using R (version 4.4.1, R Foundation for Statistical Computing, Open GNU License). An initial descriptive analysis of the data was performed, and each variable was expressed in n (%) for discrete variables and in median (with SD) for continuous variables. Receiver operating characteristic curves were generated from continuous AI scores, and the area under the curve (95% confidence interval) quantified discriminative ability. Binary performance metrics (sensitivity, specificity, positive and negative predictive values, and accuracy) were calculated at different predefined thresholds. The optimal theoretical threshold for overall diagnostic performance was determined using Youden's index (Sn + Sp − 1). Lesions were stratified by the number of positive ABCDE criteria (0, ≥1, ≥2, ≥3, ≥4, ≥5) to explore how clinical appearance influenced AI behavior and then grouped in 3 subgroups: ABCDE = 0−1, ABCDE = 2−3, and ABCDE = 4−5. Areas under the curve were computed for each subgroup with automatic orientation correction. Associations between individual ABCDE features and AI positivity (score ≥ 0.2) were tested using chi-square or Fisher's exact tests, with ORs (95% confidence interval) estimated by logistic regression. Continuous scores were compared with the Wilcoxon test. $P < .05$ was considered significant.

SUPPLEMENTARY REFERENCES

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;29:1836−42.

Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11:e047709.

Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344−9.

World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310:2191−4.

Supplementary Table S1. Study Population

Characteristics	n (%)
Sex, n (%)	
M	147 (54)
F	124 (46)
Age, y, median (minimum, maximum)	48 (18, 88)
Active smokers, n (%)	32 (12)
Significant past medical history, n (%)	21 (8)
Significant family history, n (%)	8 (3)
Obesity, n (%)	3 (1)
Fitzpatrick classification, n (%)	
II	21 (8)
III	242 (89)
IV	8 (3)
Total number of patients, n (%)	271 (100)

Abbreviations: F, female; M, male.

Supplementary Table S2. Confusion Matrix: AI Score and Pathology Analysis Performances with 0.2 Malignancy Threshold

Pathology Analysis	AI Score > 0.2 (Intermediate to High Risk)	AI Score < 0.2 (Low Risk)	Metrics
Pathology analysis: malignant	96	8	Se: 92.3%
Pathology analysis: benign	177	122	Sp: 40.8%
Metrics	PPV: 35.1%	NPV: 93.8%	Accuracy: 54.1%

Abbreviations: AI, artificial intelligence; NPV, negative predictive value; PPV, positive predictive value.
Se denotes sensitivity, and Sp denotes specificity.

Supplementary Table S3. All Lesions Final Histology

Tumor Histology	n (%)
Angiokeratoma	1 (0.2)
Botriomycoma	2 (0.5)
Benign inflammatory lesion	10 (2.5)
Basal cell carcinoma	21 (5.2)
Squamous cell carcinoma	12 (3.0)
Condyloma	2 (0.5)
Epidermic hamartoma	1 (0.2)
Hemangioma	4 (1.0)
Hisitocytofibroma	8 (2.0)
Adenomatous sebaceous hyperplasia	3 (0.7)
Lichen planus–like keratosis	5 (1.2)
Actinic keratosis	47 (11.7)
Trichilemmal cyst	3 (0.7)
Lentigo	3 (0.7)
Lichen	5 (1.2)
Bowen's disease	9 (2.2)
Melanoma	15 (3.7)
Melanosis	10 (2.5)
Molluscum	1 (0.2)
Mucocele	1 (0.2)
Naevus	174 (43.2)
Pendulum	1 (0.2)
Seborrheic keratosis	58 (14.4)
Papilloma	7 (1.7)
Total	403 (100)

Supplementary Table S4. Association between Individual ABCDE Clinical Criteria and AI Positivity (Moleanalyzer Pro Score ≥ 0.2) in the Study Cohort, Using Chi-Square Test

ABCDE Criteria	AI Score > 0.2 when Present	AI Score > 0.2 when Absent	OR (95% CI)	*P*-Value
Asymmetry	79.8%	63.3%	2.30 (1.36–3.88)	<.001
Borders irregularity	81.3%	58.2%	3.12 (1.96–4.99)	<.001
Color variegation	85.9%	57.1%	4.58 (2.71–7.74)	<.001
Diameter >0.6 mm	70.3%	63.6%	1.35 (0.88–2.07)	.166
Evolution (recent)	74.5%	58.7%	2.05 (1.34–3.13)	<.001

Abbreviations: ABCDE, Asymmetry, Borders, Color, Diameter, Evolution; AI, artificial intelligence; CI, confidence interval.
$P < .05$ was considered statistically significant.

Supplementary Table S5. Evaluation of Model Performances Based on ABCDE with AI Malignancy Thresholds Set to 0.2 and 0.5

Performance Parameters	Se (%)	Sp (%)	PPV (%)	NPV (%)	Accuracy (%)	n
AI 0.2 ABCDE = 0[1]	100	57.1	7.7	1	58.6	29
AI 0.5 ABCDE = 0	100	75.0	12.5	100	75.9	
AI 0.2 ABCDE = 1[2]	92.2	39.1	36.5	93.0	53.7	374
AI 0.5 ABCDE = 1	81.5	53.9	40.2	88.5	61.5	
AI 0.2 ABCDE = 2[3]	93.7	33.3	43.4	90.6	54.6	269
AI 0.5 ABCDE = 2	84.2	50.0	47.9	85.2	62.1	
AI 0.2 ABCDE = 3[4]	92.3	19.8	52.5	72.7	55.3	159
AI 0.5 ABCDE = 3	83.3	40.7	57.5	71.7	61.6	
AI 0.2 ABCDE = 4[5]	93.6	12.5	61.1	57.1	60.8	79
AI 0.5 ABCDE = 4	87.2	28.1	64.1	60.0	63.3	
AI 0.2 ABCDE = 5[6]	1	0	82.6	N/A	82.6	23
AI 0.5 ABCDE = 5	94.7	0	81.8	0	78.3	

Abbreviations: ABCDE, Asymmetry, Borders, Color, Diameter, Evolution; AI, artificial intelligence; CI, confidence interval; N/A, not available; NPV, negative predictive value; PPV, positive predictive value.
Se denotes sensitivity, and Sp denotes specificity.
[1]ABCDE = 0: none of the ABCDE criteria are positive.
[2]ABCDE = 1: at least 1 of the ABCDE criteria is positive.
[3]ABCDE = 2: at least 2 of the ABCDE criteria are positive.
[4]ABCDE = 3: at least 3 of the ABCDE criteria are positive.
[5]ABCDE = 4: at least 4 of the ABCDE criteria are positive.
[6]ABCDE = 5: all ABCDE criteria are positive.