TRACCIA N. 1

Il candidato illustri i documenti che compongono il bilancio unico di previsione annuale autorizzatorio.

Le fasi della procedura di affidamento dei contratti pubblici.

Quali sono le principali voci di spesa ammissibili nella rendicontazione di un progetto di ricerca finanziato con fondi pubblici (es. PNRR, PRIN, Horizon Europe)?

TRACCIA N. 2

Il candidato illustri i documenti che compongono il bilancio unico d'esercizio.

Requisiti generali e requisiti speciali per la partecipazione alle procedure di affidamento dei contratti pubblici.

Che differenza c'è tra rendicontazione a costi reali e rendicontazione a costi standard (unit cost/flat rate)? In quali casi si applica l'una o l'altra modalità?

TRACCIA N. 3

Il candidato illustri gli organi coinvolti nel processo contabile di un'università.

La programmazione degli acquisti di beni e servizi.

Quali sono i compiti del responsabile amministrativo/gestionale di un progetto di ricerca in fase di rendicontazione, e quale documentazione giustificativa deve essere conservata?

TRACCIA N. 4

Quali sono le fonti di finanziamento principali per le università.

La garanzia provvisoria e la garanzia definitiva.

Cosa si intende per 'eleggibilità della spesa' in un progetto di ricerca finanziato, e quali sono i criteri generalmente richiesti (pertinenza, tracciabilità, periodo di ammissibilità, ecc.)?

TRACCIA N. 5

Il candidato illustri gli elementi principali del ciclo passivo di un'università.

I casi di legittimo ricorso alla procedura negoziata senza bando di cui all'art. 76 del Codice dei Contratti Pubblici.
Quali conseguenze può comportare un'irregolarità riscontrata in sede di controllo/audit su un progetto di ricerca (es. rettifiche finanziarie, revoca del contributo)? Come si può prevenirla in fase di gestione amministrativa?

ORIGINAL ARTICLE

Improved detection of lentigo maligna with AI-assisted dermoscopy: A reader study in facial pigmented lesions

Abdurrahim Yilmaz, MSc,[a] Handan Merve Erol Mart, MD,[b] Burak Temelkuran, PhD,[a] and Bengu Nisa Akay, MD[b]

Background: Differentiating lentigo maligna (LM) from benign facial pigmented lesions remains difficult due to substantial clinical and dermoscopic overlap, particularly on chronically sun-damaged skin, a setting underrepresented in existing artificial intelligence (AI) studies.

Objective: To develop and evaluate a deep learning-based model for facial pigmented lesions and assess its impact on dermatology resident diagnostic performance.

Methods: In this retrospective study, 722 lesions (894 dermoscopic images) were analyzed (LM: 190; pigmented actinic keratosis: 230; solar lentigo/seborrheic keratosis: 302). Twenty percent of lesions were reserved for testing; the remainder underwent five-fold stratified cross-validation. An Xception-based convolutional neural network was trained for binary and 3-class classification. A reader study with 26 residents compared diagnostic accuracy before and after AI assistance.

Results: The model achieved a mean accuracy of 84.2% ± 2.5%, sensitivity of 90.0% ± 11.5%, and specificity of 81.9% ± 1.2%. Resident accuracy improved from 64.9% to 74.0% with AI support ($P < .0001$), with the largest gain observed in LM detection (+16.5%).

Limitations: Retrospective design, lack of multimodal clinical data, and resident-only reader study.

Conclusion: AI-assisted dermoscopy improves diagnostic performance in a challenging facial lesion setting, particularly for LM, supporting its role as an adjunct tool in clinical decision-making and dermatology training. (J Am Acad Dermatol https://doi.org/10.1016/j.jaad.2026.06.089.)

Key words: artificial intelligence; decision support system; dermoscopy; facial lesions; lentigo maligna; resident performance assessment.

INTRODUCTION

Accurate diagnosis of pigmented facial lesions remains one of the most challenging tasks in dermatology. In chronically sun-damaged skin, lentigo maligna (LM), pigmented actinic keratosis (PAK), solar lentigo (SL), and seborrheic keratosis (SK) frequently exhibit substantial clinical and dermoscopic overlap, limiting the reliability of visual assessment. This challenge is compounded by the unique histo-anatomical structure of facial skin, where the flattening

From the Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK[a]; and Department of Dermatology and Venereology, Ankara University, Ankara, Türkiye.[b]

Funding sources: Yilmaz has been funded by the President's PhD Scholarships of Imperial College London.

Patient consent: Patient consent was waived by the IRB.

IRB approval status: Reviewed and approved by Clinical Research Ethics Committee of Ankara University Faculty of Medicine; approval no: I04-388-26.

Data availability: The trained models are available at https://github.com/abdurrahimyilmaz/Facial-Lesion-Dermoscopy. The supplementary materials are available through Mendeley at https://data.mendeley.com/datasets/cdj8vwgk72.

Accepted for publication June 19, 2026.

Correspondence to: Bengu Nisa Akay, MD, Ankara Üniversitesi Tıp Fakültesi, İbni Sina Hastanesi, Akademik Yerleşke Binası, M1 Blok, Dermatoloji ABD, 06100, Altındağ, Ankara, Türkiye. E-mail: nisaakay@gmail.com.

Published online July 16, 2026.

0190-9622/$36.00

https://doi.org/10.1016/j.jaad.2026.06.089

of rete ridges gives rise to a pseudonetwork pattern, rendering classical dermoscopic criteria less specific.[1]

From a clinical perspective, accurate identification of LM is critical, given its potential to progress to invasive melanoma if left untreated. At the same time, diagnostic uncertainty leads to a substantial number of unnecessary biopsies of benign lesions, particularly in cosmetically sensitive facial areas, with implications for patient morbidity, aesthetic outcomes, and healthcare costs. This diagnostic dilemma, balancing early detection with avoidance of overtreatment, remains unresolved. The diagnostic difficulty is particularly pronounced in early lesions of LM, which often lack distinctive clinical features and may closely mimic benign lesions. Even well-established dermoscopic criteria, such as asymmetric pigmented follicular openings, rhomboidal structures, and gray dots or globules, are not entirely specific and may be observed across different entities within the differential diagnosis.[2] This substantial feature overlap highlights the intrinsic limitations of pattern-recognition-based diagnostic approaches.

CAPSULE SUMMARY

- AI applications in dermoscopy have largely focused on mixed or nonfacial datasets, with limited attention to diagnostically challenging facial lesions.
- AI-assisted dermoscopy improves diagnostic accuracy for lentigo maligna and may enhance decision-making in difficult facial lesions, particularly for less experienced clinicians.

Recent advances in dermoscopy have introduced additional criteria aimed at improving diagnostic accuracy.[3] Notably, perifollicular linear projections have been described as a novel dermoscopic feature associated with LM, representing pigmented linear structures radiating from follicular openings.[4] These structures likely correspond to early follicular invasion by atypical melanocytes and may represent an intermediate stage in the progression from asymmetric follicular pigmentation to angulated or rhomboidal structures. Although such features improve diagnostic specificity, their sensitivity remains limited.

Recent conceptual advances have challenged the conventional paradigm of dermoscopic diagnosis, which relies on the identification of specific malignant features. The "inverse approach" suggests that the absence of characteristic benign patterns, rather than the presence of individual malignant criteria, may represent a more reliable strategy for the recognition of early LM. This perspective reflects the substantial overlap of dermoscopic features among facial pigmented lesions and underscores the importance of contextual and combinatorial pattern analysis over isolated feature detection.[5]

Artificial intelligence (AI), particularly deep learning-based image analysis, offers a promising avenue to address this challenge by enabling the integration of complex, overlapping, and sometimes contradictory visual features in a reproducible manner.[6] Importantly, such systems are not strictly constrained by predefined pattern-based rules and may, therefore, capture subtle morphological relationships that are difficult to recognize through human interpretation alone.[7] While AI systems have demonstrated expert-level performance in various dermatologic classification tasks,[8-10] facial pigmented lesions represent a uniquely difficult domain due to subtle morphological differences, high inter-class similarity, and the lack of single pathognomonic features.

In this study, we develop and evaluate a deep learning-based diagnostic system for the classification of facial pigmented lesions, with particular emphasis on distinguishing LM from its major benign mimickers, including PAK and SL/SK. We then investigate the effect of AI-assisted malignancy scoring on the diagnostic performance of dermatology residents.

METHODS

This retrospective study was designed to evaluate the diagnostic performance of a deep learning model for classifying dermoscopic images of facial pigmented lesions. Ethical approval was obtained from the Clinical Research Ethics Committee of Ankara University Faculty of Medicine (Approval No: I04-388-26). The study was conducted in accordance with the principles of the Declaration of Helsinki. Due to the retrospective nature of the image analysis and the use of de-identified data, the institutional review board waived the requirement for informed consent. An overview of this study is shown in Fig 1.

Dataset collection and AI model

Dermoscopic assessment was performed using DermLite DL3N, DL4, and DL5 (3 GEN, LLC, Dana Point, CA, USA), Heine Delta 20T (Heine Optotechnik, Herrsching, Germany), and DermLite Foto II Pro light dermatoscopes in both polarized and non-polarized modes. Clinical and dermoscopic images were captured using a Nikon D7500 camera (Nikon Corp., Tokyo, Japan) attached to the dermoscopes and subsequently transferred to a computer for detailed evaluation. The dataset comprised dermoscopic images (.jpg format) categorized into 3 specific facial lesion types: LM, PAK, and SL/SK. SL and SK were

Abbreviations used:

AI: artificial intelligence
CNN: convolutional neural network
LM: lentigo maligna
PAK: pigmented actinic keratosis
SK: seborrheic keratosis
SL: solar lentigo

combined into a single class due to their well-recognized clinical, histopathological, and dermoscopic overlap on facial skin. These lesions are considered part of a continuous benign spectrum and represent the most common benign mimickers of LM in clinical practice, making their combined evaluation methodologically appropriate for the primary diagnostic task. For binary classification, lesions were grouped into 2 clinical categories: malignant (LM) and benign (PAK and SL/SK).

The study population consisted of patients from Turkiye, predominantly with Fitzpatrick skin types II-V, with the majority classified as types III-IV. Male patients were slightly more represented, comprising approximately 60% to 65% of the cohort. Histopathological confirmation was performed for all lesions except those meeting unequivocal benign diagnostic criteria. All cases of LM were histopathologically confirmed. When multiple dermoscopic images were available for a lesion, all images of sufficient technical quality were included in the dataset. Images containing significant artifacts or inadequate image quality (eg, poor focus or illumination) were excluded.

A convolutional neural network (CNN) based on the Xception[11] architecture was utilized for AI model training (Supplementary Methods, available via Mendeley at https://data.mendeley.com/datasets/cdj8vwgk72.).

Resident reader study and statistical analysis

Resident testing was conducted with 26 dermatology residents with a mean clinical experience of 2.7 years (range, 1-4 years). Each resident completed a timed, two-step diagnostic assessment in an examination-like setting, with 1 minute allotted per question. In the first step, residents classified pigmented facial lesions without any additional support. In the second step, the same cases were re-evaluated after disclosure of the AI-generated malignancy score, ranging from 0 to 100. The test set included cases from the main diagnostic categories of LM, PAK, and SL/SK. Diagnostic performance before and after AI assistance was compared using accuracy and Cohen's kappa statistics. A paired t-test was used to assess the statistical significance of the change in diagnostic accuracy following AI assistance.

For deep learning model analysis, classification performance on the hold-out test set was analyzed across individual folds. The key performance metrics reported included accuracy, sensitivity, specificity, and F1-score. Final performance is reported as the arithmetic mean, along with the standard deviation, across the 5 cross-validation folds.

RESULTS

General dataset characteristics

A total of 722 unique facial pigmented lesions encompassing 894 images (LM: 190 lesions/236 images, PAK: 230/268, SL/SK: 302/390) were evaluated (Table I). The malignant group comprised 190 LM lesions (236 images), while the benign group was represented by 230 PAK (268 images) and 302 SL/SK lesions (390 images), for a total of 532 lesions and 658 images. After allocating 20% of the lesions (144 test lesions/177 images: LM = 50, PAK = 48, SL/SK = 79) to the independent test set, the remaining patient/lesion groups (578 train + validation lesions/717 images) were distributed across 5 stratified cross-validation folds, ensuring no patient overlap between the training, validation, and test sets.

Diagnostic model performance

The Xception-based architecture reliably discriminated between LM and benign keratotic/lentiginous facial lesions. Following rigorous lesion-bound five-fold cross-validation on the held-out distinct test set, the ensemble average demonstrated robust predictive strength.

In five-fold lesion-based cross-validation (test set, $n = 177$ images; 50 malignant, 127 benign), the binary model achieved a test accuracy of 84.2% ± 2.5%, sensitivity of 90.0% ± 11.5%, and specificity of 81.9% ± 1.2% (Table I). Across the five-folds, the mean confusion matrix components were true positive = 45, false positive = 23, true negative = 104, and false negative = 5. The 3-class model achieved a mean test accuracy of 71.4% ± 0.6%, a macro-F1 score of 70.8% ± 0.9%, and a weighted-F1 score of 71.3% ± 0.7%.

Resident results

The mean baseline diagnostic accuracy of the residents was 64.8% (1686/2600; range, 50%-84%), which increased to 74.0% (1925/2600; range, 65%-89%) after access to the AI malignancy score (Fig 2 and Table II), corresponding to a mean absolute improvement of 9.2%. This improvement was highly significant on paired analysis ($P < .0001$). Across all 2600 resident-level decisions and 5200

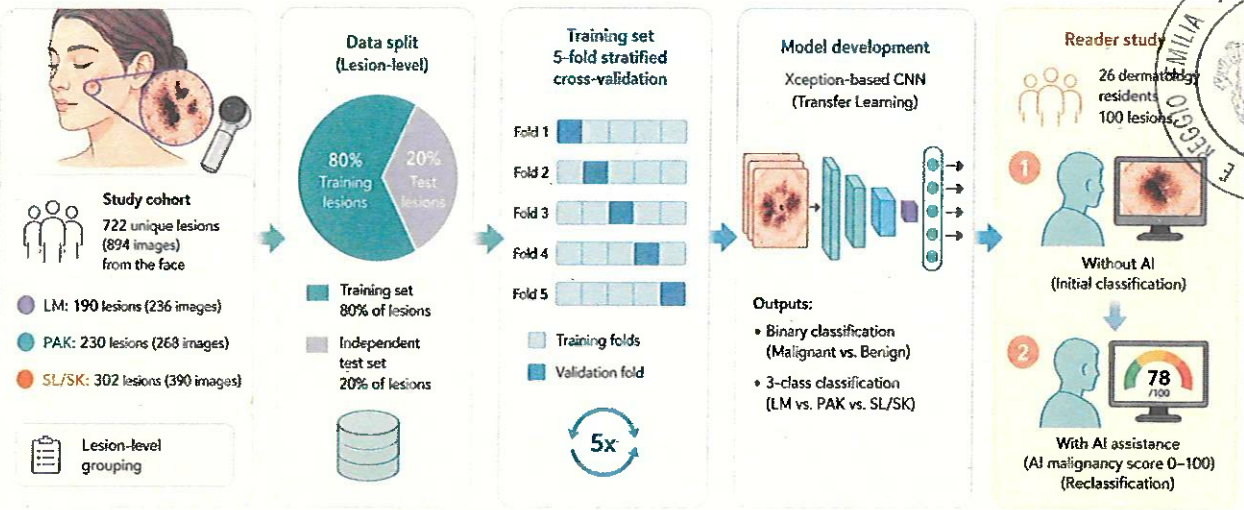

Fig 1. Overview of the study workflow. Dermoscopic images of facial pigmented lesions were collected and grouped at the lesion level to prevent data leakage. After preprocessing and lesion-level splitting, the dataset was used for training and evaluating Xception-based deep learning models for binary and 3-class classification. A reader study was then conducted to assess the effect of AI-generated malignancy scores on dermatology resident performance.

Table I. Baseline characteristics of the lesion dataset

Feature	Category	*N* (Lesion)	*N* (Image)
Total dataset		722	894
Lesion type	Lentigo Maligna (LM)	190	236
	Pigmented Actinic Keratosis (PAK)	230	268
	Solar lentigo/Seborrheic keratosis (SL/SK)	302	390
Clinical groups	Malign (LM)	190	236
	Benign (PAK + SL/SK)	532	658

total decisions, AI assistance changed incorrect to correct classifications in 302 instances, while changing correct to incorrect classifications in only 63 instances (misleading ratio of 3.74%), yielding a net gain of 239 correct diagnoses. In other words, for every case in which the AI misled a resident away from a correct answer, it corrected more than 4 previously incorrect decisions. Accuracy gains were observed across all residency years, with the greatest improvement seen among second-year residents (+11.7%), followed by third-year (+9.3%), first-year (+8.6%), and fourth-year residents (+7.6%) (Supplementary Fig 1, available via Mendeley at https://data.mendeley.com/datasets/cdj8vwgk72 and Table II). Accuracy increased from 60.6% to 69.2% among first-year residents and from 69.5% to 77.1% among fourth-year residents. Lesion-specific analysis showed that the largest benefit of AI assistance was observed for LM, where accuracy improved from 61.5% to 78.0% (+16.5%), whereas smaller gains were seen for PAK (69.8% to 75.6%, +5.8%) and SL/SK (63.5% to 68.3%, +4.8%). AI assistance also improved agreement with the reference standard, with Cohen's kappa increasing from 0.4716 to 0.6097, corresponding to a net gain of +0.1381 and a shift from moderate to substantial agreement (Table II).

DISCUSSION

The diagnosis of pigmented facial lesions remains a major challenge in dermatologic practice because LM, PAK, SL, and SK frequently share overlapping clinical and dermoscopic features. This challenge is further amplified by the distinctive histoanatomical characteristics of facial skin, in which pseudonetwork structures and follicular pigmentation patterns may reduce the specificity of conventional dermoscopic criteria. In this context, the present study evaluated a deep learning-based system for the classification of facial pigmented lesions and demonstrated that AI assistance can provide meaningful diagnostic support, particularly for the recognition of LM.

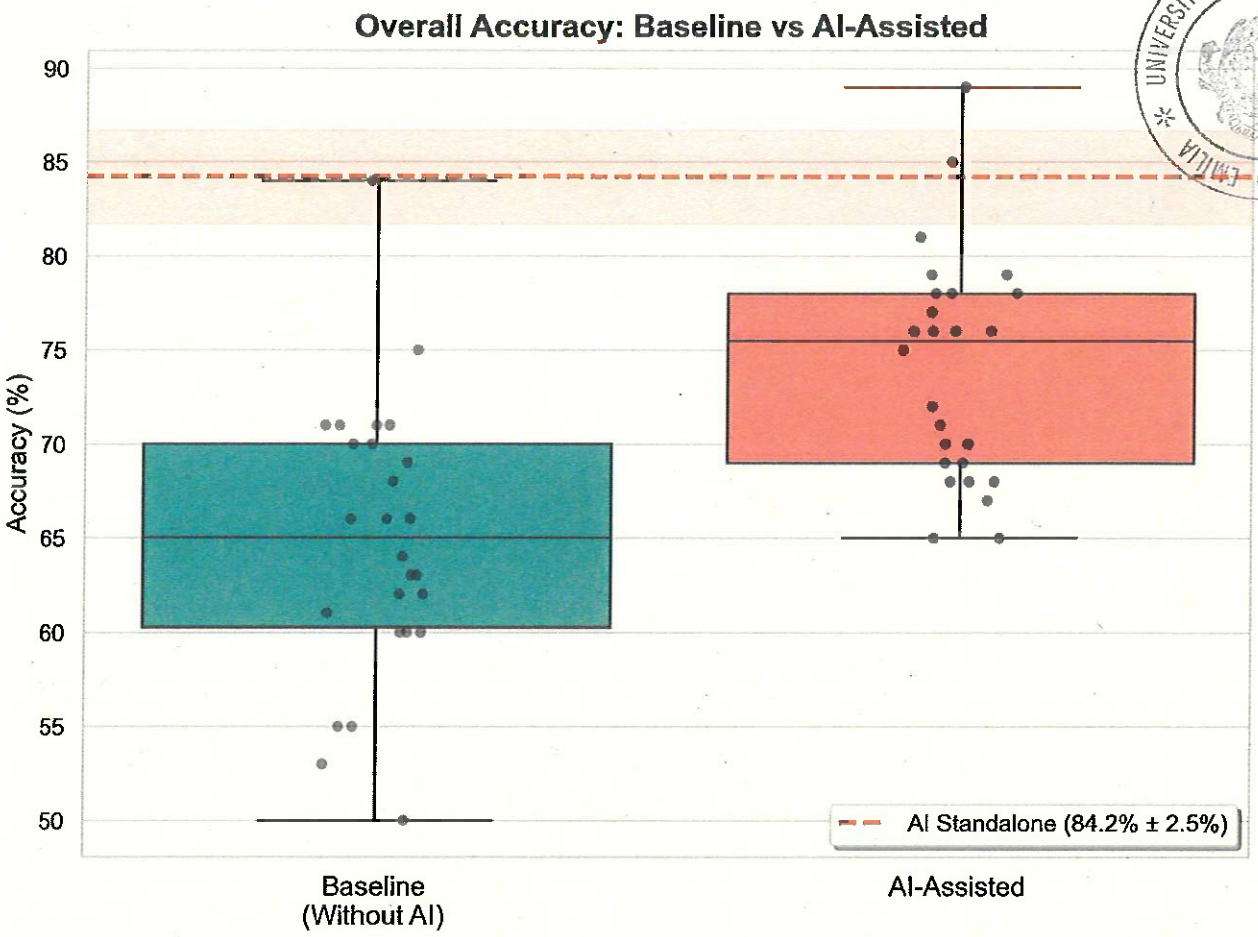

Fig 2. Reader-study results. Box-and-whisker plots with individual reader values show diagnostic accuracy before and after AI assistance. The dashed red line indicates standalone AI accuracy, and the shaded region represents the corresponding standard deviation.

The reader study provides evidence for the translational relevance of the proposed system. Across 26 dermatology residents and 2600 final and 5200 total decisions, the overall effect of AI assistance was clearly favorable, with corrected decisions substantially outweighing AI-induced errors. Nevertheless, the existence of AI-induced errors remains important and should not be overlooked. Some true LM cases were assigned relatively low malignancy scores, potentially resulting in false reassurance, whereas some benign lesions received mildly elevated scores, which may have led to unnecessary diagnostic suspicion (Supplementary Figs 1.a, 2, available via Mendeley at https://data.mendeley.com/datasets/cdj8vwgk72, and Fig 3). These observations highlight an important principle for clinical implementation: AI outputs should be interpreted as probabilistic decision-support signals rather than definitive recommendations. Effective human-AI interaction will therefore depend not only on model accuracy, but also on interface design, score calibration, user training, and clinician awareness of situations in which algorithmic outputs may be misleading.[12-14] More informative AI interfaces may further increase this benefit. Rather than providing only a malignancy score, explainable AI approaches that highlight relevant image regions or specific dermoscopic patterns may offer more interpretable support and further improve resident performance.[14] In addition, the findings indicate a positive association between residents' trust in AI output and the displayed probability of malignancy (Supplementary Fig 1.a, 1.e and 1.f, available via Mendeley at https://data.mendeley.com/datasets/cdj8vwgk72.).[15] From a clinical perspective, this finding is particularly relevant because it suggests that the model's value extends beyond isolated performance metrics and translates into measurable improvement in human diagnostic behavior.[16-18] This finding is reassuring, particularly when interpreted alongside the much larger number of corrected errors.

The lesion-specific resident results are especially notable. The largest improvement with AI assistance

Table II. Resident diagnostic performance before and after AI assistance

Metric	Without AI (Baseline)	AI-Assisted	Improvement	P-Value
AI accuracy	NA	84.2%		
Overall diagnostic accuracy of residents	64.9%	74.0%	+9.2%	<.0001
Overall Cohen's kappa of residents	0.4716	0.6097	+0.1381	<.0001
Based on lesion type				
Lentigo Maligna (LM)	61.5%	78.0%	+16.5%	
Pigmented Actinic Keratosis (PAK)	69.8%	75.6%	+5.8%	
Solar Lentigo/Seborrheic Keratosis (SL/SK)	63.5%	68.3%	+4.8%	
Based on experience				
Year 1	60.6%	69.2%	+8.6%	
Year 2	62.5%	74.2%	+11.7%	
Year 3	64.6%	73.9%	+9.3%	
Year 4	69.5%	77.1%	+7.6%	

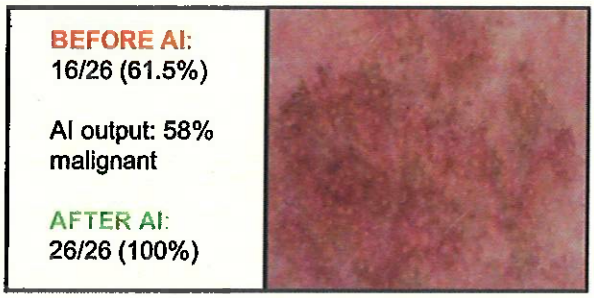

Fig 3. Lentigo maligna. The case of lentigo maligna showing one of the largest improvement in diagnostic accuracy with AI assistance. Before AI assistance, 16 of 26 residents correctly classified the lesion. After presentation of a 58% predicted probability of LM, all 26 residents classified it correctly.

was observed in LM, with diagnostic accuracy increasing from 61.5% to 78.0%. This is the most clinically meaningful category in the present study, as missed LM may progress to invasive melanoma, leading to worse prognosis and more extensive surgical treatment. These findings suggest that the AI system was particularly effective in reducing one of the most important real-world diagnostic errors in facial dermoscopy: misclassifying LM as a benign pigmented lesion, especially SL or SK. However, greater emphasis on detecting LM may also increase false-positive diagnoses and unnecessary biopsies of benign facial lesions. In the cosmetically sensitive facial setting, this may result in scarring, patient anxiety, and increased healthcare utilization. Nevertheless, this trade-off may be acceptable given the greater potential harm associated with missed melanoma.

The pattern of benefit across training levels is also noteworthy. Although all resident groups improved with AI support, the greatest gain was observed among second-year residents (+11.67%), whereas fourth-year residents showed the smallest relative gain, consistent with their higher baseline performance and reduced margin for improvement. This pattern may reflect a ceiling effect, whereby more experienced residents already perform closer to their maximum unaided level and therefore have less opportunity for measurable improvement. At the same time, the observation that all training levels benefited suggests that AI support may function both as a diagnostic aid and as an educational calibration tool. In residency settings, such systems may reinforce recognition of high-risk lesions, provide immediate probabilistic feedback, and potentially accelerate learning in diagnostically difficult domains.[19-21]

The standalone performance of the AI model also warrants consideration. The model achieved higher diagnostic accuracy than the baseline performance of residents (Accuracy: 84.2% vs 64.9%), suggesting that deep learning systems can capture diagnostically relevant patterns even in a highly challenging facial lesion setting. However, the combination of human and AI did not fully reach the standalone model performance (Accuracy: 74.0% vs 84.2%), highlighting that effective integration of AI into clinical workflows is not purely additive. This observation underscores that the clinical value of AI lies not in replacing physicians, but in complementing human decision-making. Optimizing human-AI interaction remains a critical component for translating high standalone model performance into maximal clinical benefit. The present findings support the broader concept that AI may be particularly valuable in dermatology when used as an assistive system. Our results, therefore, align with an emerging view that the most clinically useful AI systems may be those that improve physician

performance, consistency, and confidence rather than those designed solely to outperform clinicians on isolated benchmarking tasks.[17,22,23]

Several limitations should be considered when interpreting these findings and their clinical applicability. First, this was a retrospective study based on a relatively limited dataset derived from a specialized lesion category. Furthermore, the dataset originated from a single population, which may limit the generalizability of the findings to other clinical settings and patient populations. In addition, the model used dermoscopic images alone and did not incorporate clinical photographs, age, lesion history, anatomic subsite, or longitudinal change, all of which may influence real-world diagnostic reasoning. Finally, the resident study was also conducted in a time-constrained examination format and therefore does not fully capture routine clinical workflow, where biopsy thresholds, patient history, and serial follow-up also inform decision-making.

In conclusion, the present study demonstrates that deep learning-based analysis of dermoscopic facial lesions can provide clinically meaningful support for the diagnosis of LM and its principal benign mimickers. AI assistance significantly improved resident diagnostic accuracy and agreement with the reference standard, with a strongly favorable balance between corrected and induced errors. These findings support the potential role of AI-based malignancy scoring as an adjunct tool in facial dermoscopy, particularly for challenging lesions and in training environments. Future studies should focus on validation in larger and more diverse patient populations, explainable AI, prospective validation, evaluation among experienced dermatologists, external multicenter testing, integration of multimodal clinical data, and optimization of human-AI interaction to better define the role of such systems in real-world dermatologic practice. Such efforts may help bridge the gap between promising algorithmic performance and safe, effective clinical adoption.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT 5.5 for manuscript proofreading and preparation of Fig 1. Following the use of this tool, the authors carefully reviewed and edited the content as necessary and assume full responsibility for the final published article.

Abdurrahim Yilmaz has been funded by the President's PhD Scholarships of Imperial College London. The authors would like to thank the dermatology residents who participated in the reader study for their time and contribution to this work.

Conflict of interest

None disclosed.

REFERENCES

1. Lallas A, Argenziano G, Moscarella E, Longo C, Simonetti V, Zalaudek I. Diagnosis and management of facial pigmented macules. *Clin Dermatol.* 2014;32(1):94-100. https://doi.org/10.1016/j.clindermatol.2013.05.030
2. Lallas A, Tschandl P, Kyrgidis A, et al. Dermoscopic clues to differentiate facial lentigo maligna from pigmented actinic keratosis. *Br J Dermatol.* 2016;174(5):1079-1085. https://doi.org/10.1111/bjd.14355
3. Müller C, Wolber A, Lallas A, et al. Comparative and complementary diagnostic value of dermatoscopy and clinical close-up photography in skin cancer diagnosis: a study from the MILK10k dataset. *J Am Acad Dermatol.* 2026;95(1):70-77. https://doi.org/10.1016/j.jaad.2026.03.069
4. Navarrete-Dechent C, Jaimes N, Dusza SW, et al. Perifollicular linear projections: a dermatoscopic criterion for the diagnosis of lentigo maligna on the face. *J Am Acad Dermatol.* 2024;90(1):52-57. https://doi.org/10.1016/j.jaad.2023.07.1036
5. Lallas A, Lallas K, Tschandl P, et al. The dermoscopic inverse approach significantly improves the accuracy of human readers for lentigo maligna diagnosis. *J Am Acad Dermatol.* 2021;84(2):381-389. https://doi.org/10.1016/j.jaad.2020.06.085
6. Jeong HK, Park C, Henao R, Kheterpal M. Deep learning in dermatology: a systematic review of Current approaches, outcomes, and limitations. *JID Innov.* 2023;3(1):100150. https://doi.org/10.1016/j.xjidi.2022.100150
7. Tanno R, Barrett DGT, Sellergren A, et al. Collaboration between clinicians and vision–language models in radiology report generation. *Nat Med.* 2025;31(2):599-608. https://doi.org/10.1038/s41591-024-03302-1
8. 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(1):180161. https://doi.org/10.1038/sdata.2018.161
9. Yilmaz A, Yasar SP, Gencoglan G, Temelkuran B. DERM12345: a large, multisource dermatoscopic skin lesion dataset with 40 subclasses. *Sci Data.* 2024;11(1):1302. https://doi.org/10.1038/s41597-024-04104-3
10. Cartocci A, Luschi A, Tognetti L, et al. Comparative analysis of AI models for atypical pigmented facial lesion diagnosis. *Bioengineering.* 2024;11(10):1036. https://doi.org/10.3390/bioengineering11101036
11. Chollet, François. Xception: deep learning with depthwise separable convolutions. *Proc IEEE Conf Comput Vis Pattern Recognit.* 2017:1251-1258. https://doi.org/10.1109/CVPR.2017.195
12. Wolf D, Hillenhagen H, Taskin B, et al. Your other left! vision-language models fail to identify relative positions in medical images. In: Gee JC, Alexander DC, Hong J, et al., eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025.* 15964. Lecture Notes in Computer Science. Springer; 2026:691-701. https://doi.org/10.1007/978-3-032-04971-1_65
13. Gill A, Rainey C, McLaughlin L, et al. Artificial Intelligence user interface preferences in radiology: a scoping review. *J Med Imaging Radiat Sci.* 2025;56(3):101866. https://doi.org/10.1016/j.jmir.2025.101866

14. Dudley JJ, Kristensson PO. A review of user interface design for interactive machine learning. *ACM Trans Interact Intell Syst*. 2018;8(2):1-37. https://doi.org/10.1145/3185517
15. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. *JMIR AI*. 2024;3:e53207. https://doi.org/10.2196/53207
16. Azarfar G, Naimimohasses S, Rambhatla S, et al. Responsible adoption of multimodal artificial intelligence in health care: promises and challenges. *Lancet Digit Health*. 2025;7(12):100917. https://doi.org/10.1016/j.landig.2025.100917
17. Heudel PE, Crochet H, Filori Q, Bachelot T, Blay JY. Artificial intelligence in medicine: a scoping review of the risk of deskilling and loss of expertise among physicians. *ESMO Real World Data Digit Oncol*. 2026;12:100693. https://doi.org/10.1016/j.esmorw.2026.100693
18. Shekar S, Pataranutaporn P, Sarabu C, Cecchi GA, Maes P. People overtrust AI-Generated medical advice despite low accuracy. *NEJM AI*. 2025;2(6):Aloa2300015. https://doi.org/10.1056/Aloa2300015
19. Hedderich DM, Keicher M, Wiestler B, et al. AI for Doctors—A course to educate medical professionals in artificial intelligence for medical imaging. *Healthcare*. 2021;9(10):1278. https://doi.org/10.3390/healthcare9101278
20. Xu Y, Jiang Z, Ting DSW, et al. Medical education and physician training in the era of artificial intelligence. *Singapore Med J*. 2024;65(3):159-166. https://doi.org/10.4103/singaporemedj.SMJ-2023-203
21. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. 2019;5(2):e16048. https://doi.org/10.2196/16048
22. Yilmaz A, Yuceyalcin F, Varol R, et al. Resource-efficient medical vision language model for dermatology via a synthetic data generation framework. *medRxiv*. 2025. https://doi.org/10.1101/2025.05.17.25327785
23. Yao J, Li S, Liang P, Xu X, Elder D, Huang Z. Melan-Dx: a knowledge-enhanced vision-language framework improves differential diagnosis of melanocytic neoplasm pathology. *Npj Digit Med*. 2026;9(1):171. https://doi.org/10.1038/s41746-026-02357-3