Check for updates

International Journal of Dermatology

WILEY

VIEWPOINT OPEN ACCESS

AI Dermatology's Unfulfilled Promise: Why Early Melanoma Detection Remains Unverified

Tim J. Zeuner[1,2] | Titus J. Brinker[1]

[1]Division of Digital Prevention, Diagnostics and Therapy Guidance, German Cancer Research Center (DKFZ), Heidelberg, Germany | [2]Medical Faculty, University Heidelberg, Heidelberg, Germany

Correspondence: Titus J. Brinker (titus.brinker@dkfz.de)

Received: 22 April 2026 | **Revised:** 25 June 2026 | **Accepted:** 10 August 2026

Keywords: artificial intelligence | Breslow thickness | dermoscopy | diagnostic uncertainty | early detection | explainable AI | histopathology | longitudinal imaging | machine learning | melanoma

ABSTRACT

A growing number of artificial intelligence (AI) systems for melanoma diagnosis now match dermatologist-level performance. Early detection is a driving factor, yet it is rarely validated beyond aggregate diagnostic accuracy. More targeted studies are necessary so these claims become verifiable and the promise of early detection can actually be delivered to patient care.

Early detection is cited as a primary motivation in nearly all publications on artificial intelligence (AI)-assisted melanoma diagnosis. The rationale is compelling: survival rates drop substantially with increasing Breslow thickness, and detecting melanoma at its earliest stages remains the most effective way to reduce melanoma mortality. AI systems have demonstrated diagnostic performance comparable to that of dermatologists in prospective clinical studies [1]. To improve patient care, however, integrating artificial intelligence into dermatology involves more than maximizing sensitivity and specificity. Available studies provide little insight into how models perform across clinically relevant strata such as Breslow thickness or stage, and remain too diagnostically focused to establish clinical utility.

The challenge of early melanoma detection extends beyond dermatoscopy and into the current histopathological reference standard. Haggenmüller et al. [2] assessed interrater agreement among international expert pathologists reviewing histopathological whole-slide images, the highest-quality diagnostic input available. Noninvasive and pT1 melanomas showed substantial diagnostic variability among expert pathologists, with moderate interrater agreement and frequent absence of majority consensus.

These are not peripheral edge cases: Thin and noninvasive melanomas represent the majority of detections in screening settings. If expert pathologists cannot reach a reliable consensus on early-stage lesions, the implications for AI-based models for melanoma detection are serious. These systems are trained and validated on labels derived from a diagnostic process that is uncertain itself, and operate on dermoscopic or clinical images that carry far less diagnostic information. AI may therefore reproduce or amplify contemporary diagnostic uncertainty, and the claimed performance in early-stage melanoma detection warrants greater scrutiny than the current literature affords.

A recent benchmark underscored how aggregate accuracy may obscure early-stage failure. Most public datasets lack Breslow thickness or T-category labels, and ISIC challenge-winning models performed substantially worse on a benchmark enriched for in situ and thin melanomas than on standard ISIC datasets [3]. More clinically oriented evaluations are beginning to move beyond single-timepoint accuracy. PanDerm assessed sequential images of 89 melanomas and detected suspicious change at the first imaging time point in 77.5% of sequences, compared with 32.6% for clinicians [4]. This represents an important step toward evaluating AI systems in a manner that

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *International Journal of Dermatology* published by Wiley Periodicals LLC on behalf of the International Society of Dermatology.

more closely reflects clinical reality and remains one of the few approaches that explicitly address early detection. While such metrics are inherently threshold-sensitive and do not fully capture specificity trade-offs, they provide a valuable framework for moving beyond static accuracy measures.

Yet even stage-stratified or longitudinal validation addresses only part of the evidence gap. Early detection remains essential for clinically consequential melanoma, but the value of detecting every additional early lesion is not self-evident. Increased melanoma incidence, particularly for in situ and thin invasive melanomas, has often occurred without corresponding increases in mortality, raising concern that some of this increase reflects overdiagnosis rather than prevention of lethal disease [5]. For AI-assisted diagnosis, this creates central clinical tension. Sensitivity-optimized systems will improve early detection of consequential melanomas, but may also increase detection of indolent lesions that would not have become clinically relevant during a patient's lifetime.

To move beyond AI-driven early melanoma detection as a recurring headline, future work must align more closely with the realities of diagnostic uncertainty and clinical decision-making [6]. Against the background of effective immunotherapy and targeted therapy, the incremental benefit of detecting ever-earlier lesions should be demonstrated rather than assumed. Future studies should combine stage-stratified validation, especially for in situ and thin melanomas, with longitudinal datasets and clinically meaningful endpoints such as melanoma-specific mortality, unnecessary procedures, or surveillance burden. Without such evidence, continued improvements in aggregate diagnostic accuracy risk overstating progress while leaving the central promise of AI-driven early detection unfulfilled.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Funding

The authors have nothing to report.

Conflicts of Interest

T.J.B. is the owner of Smart Health Heidelberg GmbH, a software company which developed the first teledermatology app in Germany ("AppDoc" https://appdoc.de) and received honoraria from Novartis, Roche, Merck and HEINE Optotechnik. The other authors declare no competing interests.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

1. S. Laiouar-Pedari, A. Kühn, C. Wies, et al., "Prospective Evidence on Artificial Intelligence–Assisted Melanoma Diagnostics: A Systematic Review and Meta-Analysis," *JAMA Dermatology* 162 (2026): 478–487, https://doi.org/10.1001/jamadermatol.2026.0217.

2. S. Haggenmüller, C. Wies, J. Abels, et al., "Discordance, Accuracy and Reproducibility Study of Pathologists' Diagnosis of Melanoma and Melanocytic Tumors," *Nature Communications* 16, no. 1 (2025): 789, https://doi.org/10.1038/s41467-025-56160-x.

3. A. Dzieniszewska, P. Garbat, P. Pietkiewicz, and R. Piramidowicz, "Early-Stage Melanoma Benchmark Dataset," *Cancers* 17, no. 15 (2025): 2476, https://doi.org/10.3390/cancers17152476.

4. S. Yan, Z. Yu, C. Primiero, et al., "A Multimodal Vision Foundation Model for Clinical Dermatology," *Nature Medicine* 31, no. 8 (2025): 2691–2702, https://doi.org/10.1038/s41591-025-03747-y.

5. M. F. Bjørch, E. G. Gram, and J. B. Brodersen, "Overdiagnosis in Malignant Melanoma: A Scoping Review," *BMJ Evidence-Based Medicine* 29, no. 1 (2024): 17–28, https://doi.org/10.1136/bmjebm-2023-112341.

6. T. Chanda, C. Wies, F. Schramm, et al., "Large Language Models Improve Physician Accuracy but Lead to False Reliance," *arXiv* preprint arXiv:2608.00817 (2026), https://doi.org/10.48550/arXiv.2608.00817.

13654632, 2026, S1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ijd.70636 by Deutsches Krebsforschungszentrum (DKFZ), Wiley Online Library on [13/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License