

## EDITORIAL OPEN



# AI-based skin cancer detection: the balance between access and overutilization

Gregoor et al. evaluated the healthcare implications and costs of an AI-enabled mobile health app for skin cancer detection, involving 18,960 beneficiaries of a Netherlands insurer. They report a 32% increase in claims for premalignant and malignant skin lesions among app users, largely attributed to benign skin lesions and leading to higher annual costs for app users (€64.97) compared to controls (€43.09). Cost-effectiveness analysis showed a comparable cost to dermatologist-based diagnosis alone. This editorial emphasizes the balance in AI-based dermatology between increased access and increased false positives resulting in overutilization. We suggest refining the diagnostic schemas with new referral pathways to capitalize on potential savings. We also discuss the importance of econometric analysis to evaluate the adoption of new technologies, as well as adapting payment models to mitigate the risk of overutilization inherent in AI-based diagnostics such as skin cancer detection.

*npj Digital Medicine* (2023)6:147; <https://doi.org/10.1038/s41746-023-00900-0>

The use of AI for medical diagnosis has found an early home at scale in skin cancer. The complex process of diagnosis can involve integrating data on a patient's symptoms and history, physical exam, lab values, and imaging studies. AI tools, including machine learning and deep learning algorithms, can learn from and efficiently process large volumes of data. Researchers have used AI-based tools to aid in the process of diagnosis in various different contexts, including the detection of diseases of the skin<sup>1</sup>, liver<sup>2</sup>, heart<sup>3</sup>, and other organs<sup>4</sup>. Moreover, other tools are made to interface directly with patients and influence their care.

Gregoor et al. analyze data from the pilot of a mobile health app that used an AI algorithm for skin cancer detection<sup>5</sup>. In 2019, a Dutch health insurance company offered 2.2 million adults free access to this app. Gregoor et al. matched 18,960 users who completed at least one successful assessment with the app to controls who did not use the app. They reported a 32% increase in claims for premalignant and malignant skin lesions among app users compared to non-app users. App users had twice as many biopsies and excisions matched as well as four times (5.9%) the claims for benign skin tumors and nevi compared with controls (1.7%). App users also had fewer claims for malignant skin lesions than controls. The increased benign claims and fewer malignant claims resulted in higher total annual costs for app users (€64.97) vs. controls (€43.09); costs per individual claim for malignant lesions were also higher for app users (€613.36) vs. controls (€520.05). Estimates on the cost of capturing one additional premalignant and malignant skin lesions via the app ranged from €2657 to €488.

## COMPARING AI-BASED SKIN CANCER DIAGNOSIS WITH CONVENTIONAL DERMATOLOGY

This study is noteworthy in its effort to characterize the healthcare implications of dermatology AI-based diagnosis beyond the existing literature on diagnostic accuracy alone—i.e., cost, cost-effectiveness, and utilization. The reasons for overutilization enabled by this app are manifold. The app performed as expected from previous reports by the developers—sensitivity of 87–95% and specificity of 70–78%<sup>6,7</sup>. The deployment of this AI skin app at a broad scale shows the real-world costs of more false positives (benign lesion claims) and fewer true positives

(malignant lesion claims) compared to the management of non-app users. More false positives and fewer true positives compared with conventional care can take an emotional and financial toll on patients and the healthcare system.

The overall cost-effectiveness of the screening may be comparable to that of a dermatologist. A recent study in the US found that the cost of detecting an additional skin premalignancy or malignancy through total body exams was \$2346<sup>8</sup>. Depending on the assumptions of these calculations, the skin app performed at a comparable cost per new positive identification. In context, increased total costs per app user at a comparable cost-benefit ratio suggests that the app users are enjoying more of the “benefits”—i.e., they had more skin lesions diagnosed than non-app users, likely due to increased access. This supports using AI skin apps insofar as access is the limiting determinant of diagnosis.

In addition, the algorithm was preset to flag premalignancies like actinic keratosis, a dysplastic but inherently benign lesion, as high risk. Actinic keratosis is a benign lesion on the same spectrum as squamous cell carcinoma, but with a low risk (0.1%) of malignant transformation<sup>9</sup>. Higher scrutiny in flagging these findings for review/claims should be utilized to triage for more cost-effective care. This may require more nuanced classification schemas as high risk, low risk, or moderate risk. Moderate-risk categories could then be evaluated and triaged by dermatologists and advanced practice providers, with lower reimbursement or diagnostic priority to save costs.

## ADDRESSING OVERUTILIZATION

These data fall along similar findings across other areas of digital health, with the benefit of increased access and timeliness balanced by the risk of overutilization. For example, unique aspects of telehealth, such as more efficient triage and decreased emergency department utilization, suggest a potential for cost-effectiveness, paralleling the benefits and risks observed in dermatology AI. However, in another area of digital health, telehealth implementation in primary care during COVID-19 may have led to overutilization, with some studies suggesting that telehealth visits were used as additions rather than substitutes to in-person visits<sup>10</sup>. A more direct parallel to skin AI may be radiology AI, wherein increased imaging given ease of AI-enabled use led to increased invasive testing and follow-up due to false positives<sup>11</sup>.




As AI implementation will inevitably continue to increase, a few different strategies will be necessary to address the conundrum of

overutilization. Primarily, the use of AI algorithms should be rationalized before implementation. When appropriate, AI algorithms should be evaluated against traditional methods using econometric analysis and pilot studies such as that of Gregoor et al. The rate and cost of true positive identification should be assessed against the rate of false positives and the performance of traditional care. On a systems level, this will require regulators and administrators to establish guidelines for the responsible use of AI diagnostics. The V3 framework for biometric monitoring can be adapted for AI diagnostics more broadly; it involves verification of the technology with preset criteria, analytical validation of the algorithm, and clinical validation in a real-world target context<sup>12</sup>.

Payment and incentive models will need to be adapted as well. The current per-use reimbursement models, while feasible for early AI products, may result in the overuse of AI, analogous to experiences with traditional medical devices. In the pursuit of value-based care, reimbursement should incorporate outcomes instead of volume. In addition, when piloting new AI devices like the skin cancer app in this study, payers should utilize advanced market commitments and time-limited reimbursements for new AI applications. Such an approach can more sensitively control the adoption of AI technologies and mitigate risks of overuse<sup>13</sup>.

As AI for clinical diagnostics moves more broadly into the implementation stage, the threat of overutilization should be anticipated. Rising to this challenge will require adjustment to payment models and evidence-based stewardship.

Received: 19 June 2023; Accepted: 7 August 2023;  
Published online: 15 August 2023

Kaushik P. Venkatesh <sup>1</sup>✉, Marium Raza <sup>1</sup> and Joseph Kvedar <sup>1</sup>  
<sup>1</sup>Harvard Medical School, Boston, MA, USA.  
✉email: kaushik\_venkatesh@hms.harvard.edu

## REFERENCES

- Kumar, Y., Koul, A., Singla, R. & Ijaz, M. F. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J. Ambient Intell. Humaniz Comput.* **14**, 8459–8486 (2023).
- Lin, R. H. An intelligent model for liver disease diagnosis. *Artif. Intell. Med.* **47**, 53–62 (2009).
- Karatzia, L., Aung, N. & Aksentijevic, D. Artificial intelligence in cardiology: hope for the future and power for the present. *Front. Cardiovasc. Med.* **9**, 945726 (2022).
- Richens, J. G. & Buchard, A. Artificial intelligence for medical diagnosis. In: *Artificial Intelligence in Medicine* (eds Lidströmer, N. & Ashrafian, H.) 181–201 (Springer International Publishing). [https://doi.org/10.1007/978-3-030-64573-1\\_29](https://doi.org/10.1007/978-3-030-64573-1_29) (2022).
- Smak Gregoor, A. M. et al. An artificial intelligence based app for skin cancer detection evaluated in a population based setting. *NPJ Digit. Med.* **6**, 90 (2023).
- Udrea, A. et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J. Eur. Acad. Dermatol. Venereol.* **34**, 648–655 (2020).
- Sangers, T. et al. Validation of a market-approved artificial intelligence mobile health app for skin cancer screening: a prospective multicenter diagnostic accuracy study. *Dermatology* **238**, 649–656 (2022).
- Matsumoto, M. et al. Estimating the cost of skin cancer detection by dermatology providers in a large health care system. *J. Am. Acad. Dermatol.* **78**, 701–709.e1 (2018).
- Reinehr, C. P. H. & Bakos, R. M. Actinic keratoses: review of clinical, dermoscopic, and therapeutic aspects. *Bras. Dermatol.* **94**, 637–657 (2019).
- Venkatesh, K. P., Raza, M. M. & Kvedar, J. Has increased telehealth access during COVID-19 led to over-utilization of primary care? *NPJ Digit. Med.* **5**, 178 (2022).
- Hendrix, N., Veenstra, D. L., Cheng, M., Anderson, N. C. & Verguet, S. Assessing the economic value of clinical artificial intelligence: challenges and opportunities. *Value Health* **25**, 331–339 (2022).
- Goldsock, J. C. et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit. Med.* **3**, 55 (2020).
- Venkatesh, K. P., Raza, M. M., Diao, J. A. & Kvedar, J. C. Leveraging reimbursement strategies to guide value-based adoption and utilization of medical AI. *NPJ Digit. Med.* **5**, 112 (2022).

## AUTHOR CONTRIBUTIONS

The first draft was written by K.P.V. M.R. and J.K. provided critical revisions and approved the final draft.

## COMPETING INTERESTS

J.K. is the Editor-in-Chief of *npj Digital Medicine*. The other authors declare no competing interests.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023