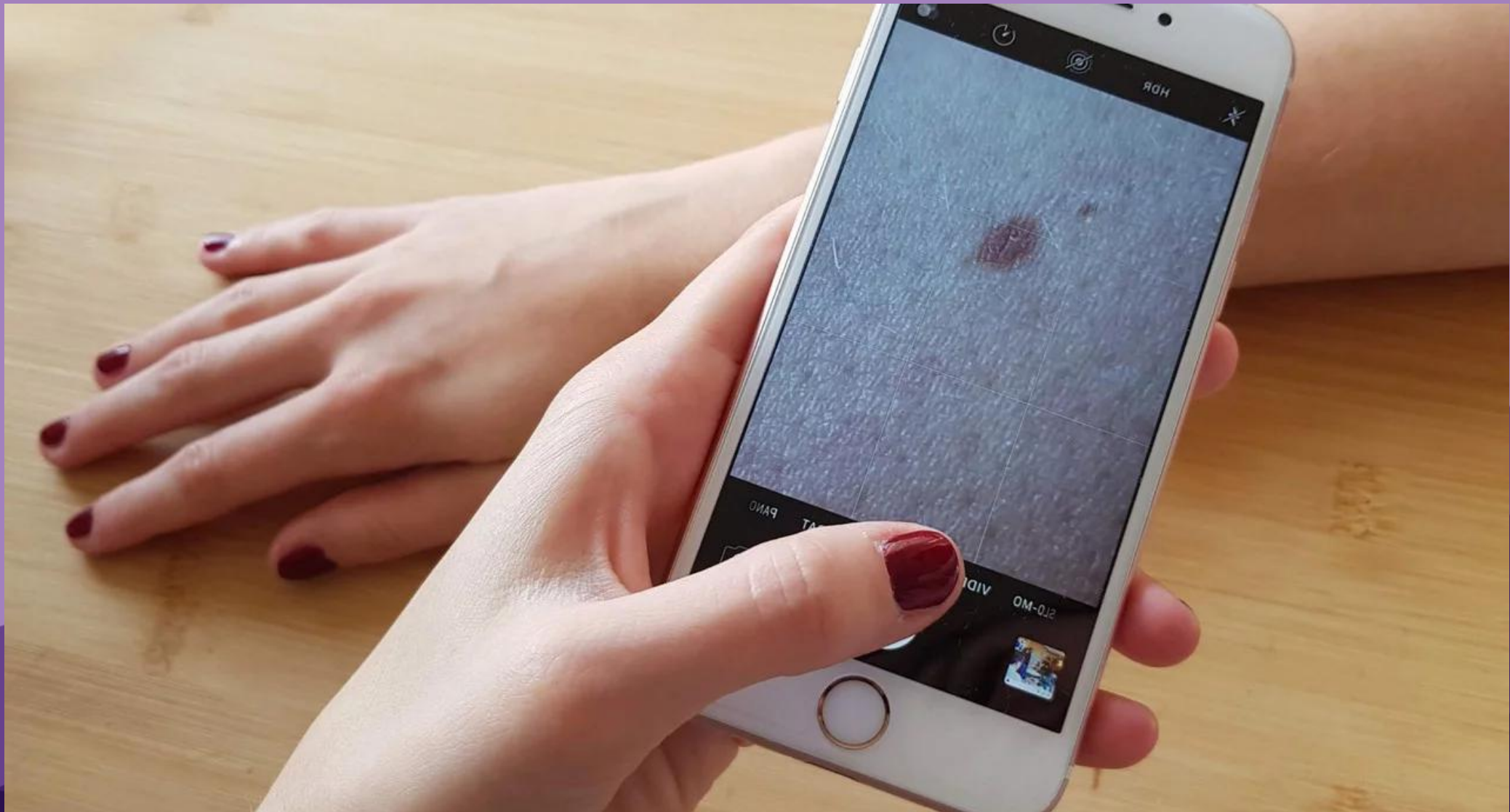




Quelle: https://sozialministerium.baden-wuerttemberg.de/fileadmin/_processed_/1/d/csm_230811_sKIn_Hautscreening_2434x1310_732a563930.jpg

Künstliche Intelligenz in der Dermatologie: Potenziale und Grenzen bei der Hautkrebsdiagnostik

18.3 Evidence-Based Practice und Evaluation

Prüferin: Andrea Jessen

Sandra Konschuh (36388)

Relevanz des Themas

Hohe und steigende Krankheitslast durch Hautkrebs

- Eine der häufigsten Krebsarten in Deutschland
- Die Zahl der Hautkrebsdiagnosen ist seit 2005 stark angestiegen

Bedeutung der Früherkennung für Heilung und Ressourcen

- Frühzeitige Erkennung verbessert Heilungschancen erheblich
- Viele können durch kleine Operationen geheilt werden
- Vermeidet teure Therapien im Spätstadium

Effizienzsteigerung und gesundheitsökonomischer Nutzen

- Steigende Erkrankungszahlen führen zu höheren Kosten und mehr stationären Behandlungen -> 2023: 116,9 Tausend
- KI-basierte Systeme optimieren Arbeitsabläufe und Ressourceneinsatz
- KI trägt zur Kostenreduktion und Vermeidung unnötiger Eingriffe bei

PICO-Frage:

Kann eine KI-basierte Software bei erwachsenen Patienten mit Verdacht auf Hautkrebs in der klinischen Praxis Hautkrebs zuverlässiger erkennen als ein erfahrener Dermatologe?

Population	Erwachsene Patienten mit Verdacht auf Hautkrebs
Intervention	Der Einsatz einer KI-basierten Software zur bildgestützten Hautkrebsdiagnose
Comparison	Diagnostik durch erfahrene Dermatologen
Outcome	Diagnostische Genauigkeit (Sensitivität, Spezifität), Sicherheit der Diagnose, ggf. Vertrauen in die Diagnose

Literatur-Auswahlkriterien

Einschlusskriterien	• Erwachsene Patienten mit Hautkrebs oder Verdacht darauf • Meta-Analysen, systematische Reviews, RCTs, Originalstudien • Peer-Reviewed, zitierte Artikel • Studien mit direktem Bezug zu KI-gestützter Hautkrebsdiagnostik (Apps, Algorithmen) • Sprache: Deutsch, Englisch
Ausschlusskriterien	• Studien zu anderen Erkrankungen als Hautkrebs • Studien ohne klinischen Bezug (z. B. rein technische KI-Entwicklung ohne Anwendung) • Kinder und Jugendliche • Nicht-peer-reviewte Quellen
Filter	• Veröffentlichungen der letzten 5 Jahre

Datenbasis

PubMed:

- 20.05.2025
- Suchbegriff: ai in dermatology comparison dermatologist
- Ergebnisse: 17
- Filter 5 Jahre

Google Scholar:

- 20.05.2025
- Suchbegriff: ai in dermatology comparison dermatologist skin cancer
- Ergebnisse: ca. 21.200
- Filter 5 Jahre

Ergebnis nach dem CRAAP-Kriterium

Currency (Aktualität):

- April 2025 veröffentlicht
- Basiert auf aktuellen KI-Methoden + aktuelle KI DEXI
- Aktuelle Herausforderungen der dermatologischen Diagnostik

Relevance (Bezug):

Population	Erwachsene Patienten mit Verdacht auf Hautkrebs	Erwachsene mit hohem Risiko auf Hautkrebs
Intervention	Der Einsatz einer KI-basierten Software zur bildgestützten Hautkrebsdiagnose	KI-System DEXI (Dermoscopy EXplainable Intelligence)
Comparison	Diagnostik durch erfahrene Dermatologen	Dermatologe mit 25 Jahren Erfahrung
Outcome	Diagnostische Genauigkeit (Sensitivität, Spezifität), Sicherheit der Diagnose, ggf. Vertrauen in die Diagnose	• Diagnostische Genauigkeit: ▪ Sensitivität ▪ Spezifität ▪ AUC • Bilddatenbank mit 1047 mit dermatologischen Bildern, wovon 302 spezielle Arten sind und die klinische Realität abbilden

Authority (Reputation):

Autoren:

Lior Mevorach, Giovanni Pellacani und Carmen Cantisani:

- Dermatologie-Einheit, Abteilung für Klinische Innere Medizin, Anästhesiologische und Kardiovaskuläre Wissenschaften, Sapienza Universität Rom

Alessio Farcomeni

- Fakultät für Wirtschaftswissenschaften, Universität Tor Vergata Rom

Herausgeber:

- Diagnostics (MDPI) ist ein peer-reviewtes Fachjournal, gelistet in PubMed, Scopus etc.

Accuracy (Sorgfalt):

- 1.047 dermatoskopische Bilder, alle histologisch bestätigt
- Klare Ergebnisdarstellung: Sensitivität, Spezifität, AUC-Wert, ROC-Kurven
- Alle Aussagen und Erklärungen mit Literatur belegt
- Peer-reviewed Veröffentlichung in Diagnostics (MDPI), indexiert in PubMed, Scopus

Purpose (Zweck):

- **Ziel:** objektiver Leistungsvergleich von KI, Dermatologe und Medizinstudent
- Keine kommerzielle Absicht, keine Werbung oder Produktplatzierung (in der Studie ausdrücklich erwähnt)
- **Inhalt:**
 - Faktenbasiert
 - Meinung nur im Diskussionsteil
 - Kritische Reflexion von Stärken & Schwächen des KI-Systems
- Keine ideologischen oder finanziellen Interessen offengelegt



Article

A Comparison of Skin Lesions' Diagnoses Between AI-Based Image Classification, an Expert Dermatologist, and a Non-Expert

Lior Mevorach ¹, Alessio Farcomeni ², Giovanni Pellacani ¹ and Carmen Cantisani ^{1,*}

¹ Dermatology Unit, Department of Clinical Internal Anesthesiological and Cardiovascular Sciences, "Sapienza" University of Rome, 00161 Rome, Italy; dermatologysapienza@gmail.com (L.M.); pellacani.giovanni@uniroma1.it (G.P.)

² Faculty of Economics, Tor Vergata University of Rome, 00133 Roma, Italy

* Correspondence: carmen.cantisani@uniroma1.it; Tel.: +39-06499769993

Abstract: Background/Objectives: This study aims to evaluate and compare the diagnostic accuracy of skin lesion classification among three different classifiers: AI-based image classification, an expert dermatologist, and a non-expert. Given the rising interest in artificial intelligence (AI) within dermatology, it is crucial to assess its performance against human expertise to determine its viability as a reliable diagnostic tool. **Methods:** This reader study utilized a set of pre-labeled skin lesion images, which were assessed by an AI-based image classification system, an expert dermatologist, and a non-expert. The accuracy of each classifier was measured and compared against the ground truth labels.

Bewertung der Studie



Stärken

- Verwendung von realen klinischen Bildern mit histologischer Bestätigung
- Drei-Vergleich: KI, erfahrener Dermatologe, medizinischer Laie
- Transparent dokumentierte Ergebnisse (AUC, Sn, Sp)
- KI (DEXI) zeigt höchste diagnostische Genauigkeit
- Ethische Perspektive miteinbezogen
- Keine Interessenkonflikte, gute Nachvollziehbarkeit

Schwächen



- Nur ein Dermatologe beteiligt → geringe Vergleichsstärke
- Nur ein KI-System getestet (DEXI) → keine externe Validierung
- Datensatz aus nur einem Land (Italien) → eingeschränkte Generalisierbarkeit
- Verschiedene Hauttypen wurden nicht betrachtet

Fazit

- KI (DEXI) übertraf den Dermatologen in der Diagnosegenauigkeit
- Hohe methodische Qualität, aber begrenzte externe Validität
- Die Studie liefert wertvolle Evidenz, dass KI-Systeme zukünftig eine ergänzende Rolle in der Hautkrebserkennung spielen könnten
- Weitere Studien mit größeren Stichproben und mehreren Ärzten/KI-Systemen sind notwendig

Literaturverzeichnis:

Grobe, T./ Vanella, P./ Szecsenyi, J. (2025): Arztreport 2025: Hautkrebs – Versorgung und Trends; BARMER Institut für Gesundheitssystemforschung, Berlin, S. 10-27.

Fraunhofer (2022): Smartphone-App und KI-Software beschleunigen Erkennung von Hautkrebs; URL: <https://www.fraunhofer.de/de/presse/presseinformationen/2022/februar-2022/smartphone-app-und-ki-software-beschleunigen-erkennung-von-hautkrebs.html> [Stand: 20.05.2025].

Rinderknecht, F.-A./ Nwandu, L./ Daneshjou, R./ Lester, J. (2024): The Impact of Artificial Intelligence on Health Equity in Dermatology in: Current Dermatology Reports (2024), Springer, S. 148-155.

Das Statistische Bundesamt (Destatis) (2025): KORREKTUR: Zahl der stationären Hautkrebsbehandlungen binnen 20 Jahren um 87,5 % gestiegen; URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/Zahl-der-Woche/2025/PD25_20_p002.html?templateQueryString=wiesbaden&templateQueryString=wiesbaden [Stand: 20.05.2025].

Mevorach, L./ Pellacani, G./ Cantisani, C./ Farcomeni, A. (2025): A Comparison of Skin Lesions' Diagnoses Between AI-Based Image Classification, an Expert Dermatologist, and a Non-Expert; MDPI.

Anhang

	Yes	No	Unclear	Not applicable
1. Was a consecutive or random sample of patients enrolled?	☒	☐	☐	☐
2. Was a case control design avoided?	☒	☐	☐	☐
3. Did the study avoid inappropriate exclusions?	☒	☐	☐	☐
4. Were the index test results interpreted without knowledge of the results of the reference standard?	☒	☐	☐	☐
5. If a threshold was used, was it pre-specified?	☐	☐	☒	☐
6. Is the reference standard likely to correctly classify the target condition?	☒	☐	☐	☐
7. Were the reference standard results interpreted without knowledge of the results of the index test?	☒	☐	☐	☐
8. Was there an appropriate interval between index test and reference standard?	☒	☐	☐	☐
9. Did all patients receive the same reference standard?	☒	☐	☐	☐
10. Were all patients included in the analysis?	☒	☐	☐	☐