

Softwarebasierte Fehlklassifikationen und Detektionslücken

In der dermatologischen Praxis werden Assistenzsysteme, die auf künstlicher Intelligenz (KI) basieren und zur Analyse dermatoskopischer Bilder eingesetzt werden, immer wichtiger¹. Diese Systeme versprechen eine objektive Risikoeinschätzung und sie unterstützen die Früherkennung von Hauttumoren. Die Diagnoseleistung der Assistenzsysteme entspricht derjenigen von Dermatolog:innen oder übertrifft sie sogar²⁻⁴. Der klinische Alltag zeigt jedoch, dass diese Systeme nicht immer fehlerfrei arbeiten⁵.

Autor:innen | cand. med. Joel Wiederkehr, Dr. med. Adamantia Milia-Argyti, Dr. med. C. Bettina Rümmelein

● Der Markt für KI-gestützte Bildanalyse in der Dermatologie teilt sich heute im Wesentlichen in zwei Anwendungsbereiche, die sich bezüglich Zielgruppe, Bildqualität und diagnostischer Präzision fundamental unterscheiden⁶:

- **App-basierte Lösungen für die Allgemeinbevölkerung:** Diese Anwendungen sind für Smartphone-Kameras konzipiert und richten sich direkt an Laien zur eigenständigen Erstüberprüfung von Muttermalen. Obwohl diese Apps teilweise als Medizinprodukte zugelassen sind, ist ihre diagnostische Genauigkeit deutlich geringer als diejenige von klinischen Untersuchungen durch Dermatolog:innen und professionellen dermatoskopischen Bildanalysen⁷. Die Gründe dafür sind variierende Lichtverhältnisse, mangelnde Standardisierung und fehlende Vergrößerung.
- **Professionelle, dermatoskopische KI-Systeme:** Diese Softwarelösungen sind direkt an die digitale Dermatoskopie in der fachärztlichen Praxis gekoppelt. Sie arbeiten unter

standardisierten Bedingungen (Polarisationsfilter, definierter Abstand, hochauflösende Makro-Optik) und nutzen Algorithmen, die klinisch validiert wurden⁸.

Genau wie Menschen machen auch Assistenzsysteme Fehler. Bei den folgenden drei Fallbeispielen wurden jeweils professionelle, praxisbasierte KI-Systeme eingesetzt, die jedoch risikoreiche Läsionen nicht erkannten. Die Fehlklassifikation als «unauffällig» erfolgte nicht auf der Basis qualitativ minderwertiger Smartphone-Bilder, sondern trotz des Einsatzes hochspezialisierter Medizintechnik. Die Fälle zeigen eindrücklich, dass der Blick und die Erfahrung von Ärzt:innen noch nicht durch Technologie ersetzbar sind. Es bleibt relevant, auch zertifizierte Systeme im klinischen Alltag stets kritisch zu hinterfragen. →



cand. med. Joel Wiederkehr
Hautwerk AG
Maneggstrasse 17
CH-8041 Zürich
klinik@hautwerk.ch



Dr. med. Adamantia Milia-Argyti
Hautwerk AG
Maneggstrasse 17
CH-8041 Zürich
milia@hautwerk.ch

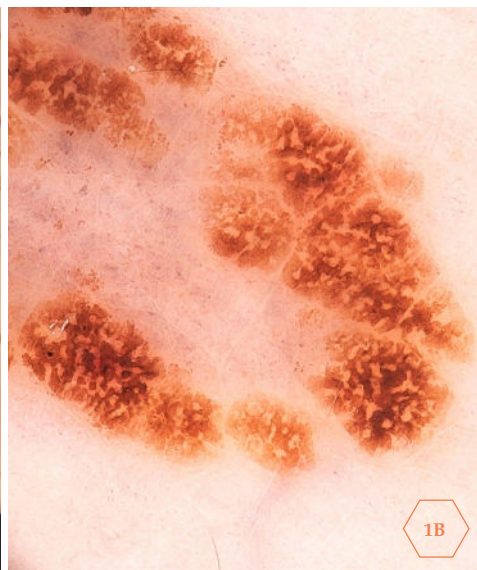


Dr. med. C. Bettina Rümmelein
Ärztliche Leitung
Hautwerk AG
Maneggstrasse 17
CH-8041 Zürich
ruemmelein@hautwerk.ch

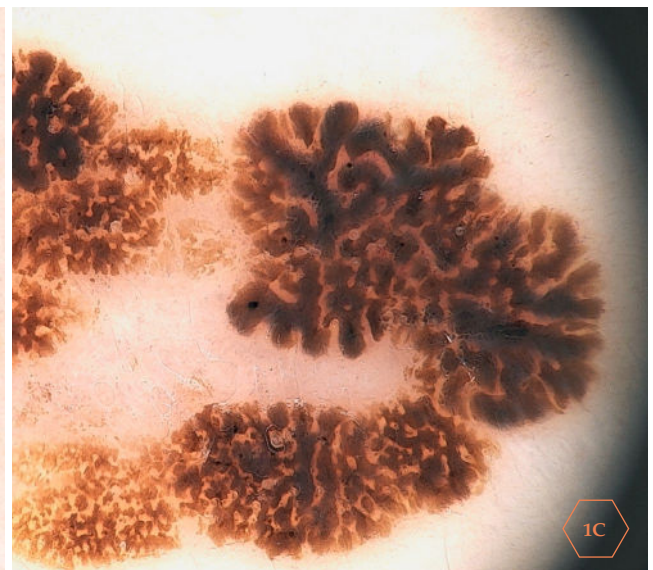
[ABB. 1A] Suspekte Läsion an der linken Flanke



[ABB. 1B] Auflichtmikroskopische Aufnahme der Läsion vor einem Jahr



[ABB. 1C] Aktuelle Aufnahme der gleichen Stelle: Deutlich zu sehen ist die Progression innerhalb eines Jahres.



Klinischer Fall 1

Ein 65-jähriger Patient mit Hauttyp II, ohne persönliche oder familiäre Melanomanamnese, stellte sich im Rahmen der jährlichen Hautkrebsvorsorge in unserer Sprechstunde vor. Bei der Ganzkörperuntersuchung mittels digitaler Dermatoskopie fiel eine klinisch suspekte Läsion an der linken Flanke auf, mit deutlicher Progredienz im Vergleich zur Erstaufnahme vor einem Jahr [ABB. 1A–C].

Zur Unterstützung wurde die integrierte KI-Analyse durchgeführt. Überraschenderweise klassifizierte der Algorithmus das Bild als «unauffällig» (AI-Score von 5%) und als nicht melanozytär. Aufgrund des dermatoskopischen Verdachts auf ein Melanom wurde dem Patienten dennoch die umgehende Exzision geraten. Durch die diskrepante, verharmlosende Bewertung des KI-Systems verunsichert, lehnte der Patient den sofortigen Eingriff jedoch zunächst ab, um andernorts eine Zweitmeinung einzuholen. Dieses Beispiel demonstriert eindrücklich, dass diskrepante KI-Befunde nicht nur diagnostische, sondern auch psychologische Risiken bergen, indem sie die Compliance der Patient:innen und damit den rechtzeitigen Therapiebeginn gefährden.

Klinischer Fall 2

Der zweite Fall betrifft einen 45-jährigen Patienten mit Hauttyp II, ebenfalls ohne Melanom in der persönlichen Vorgeschichte oder Familie, der sich zum regelmässigen Hauttumorscreening vorstellte. Hierbei fiel eine Veränderung an der rechten Flanke ins Auge, die der Patient zuvor noch nicht bemerkt hatte [ABB. 2A]. Die Läsion präsentierte sich im zweimonatigen Verlauf ohne wesentliche Dynamik [ABB. 2B und C]. Dennoch zeigte sie auflichtmikroskopische Merkmale wie ein asymmetrisches Pigmentnetzwerk, die Zweifel an der Benignität aufkommen liessen.

Die zur Absicherung hinzugezogene KI-Software bewertete die dermatoskopische Aufnahme ebenfalls als «unauffällig» (AI-Score von 5%). Aufgrund der klinischen Atypien wurde eine

Totalexzision zum Ausschluss einer Dysplasie oder Malignität empfohlen. Histologisch wurde die Diagnose eines kutanen Melanoma in situ vom Typ des superfiziell spreitenden Melanoms gesichert. Durch das rechtzeitige ärztliche Einschreiten konnte die gesamte Läsion in toto entfernt werden.

Klinischer Fall 3

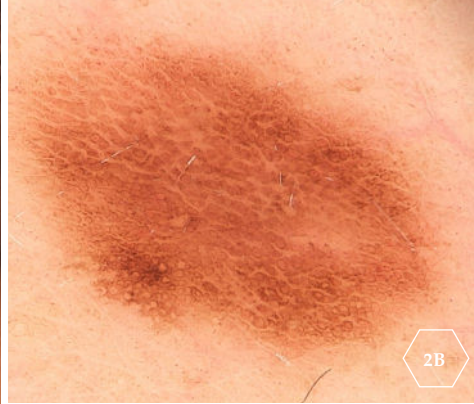
Eine 33-jährige Patientin stellte sich vor, um ein Tattoo mittels Lasertherapie entfernen zu lassen. Im Rahmen der klinischen Voruntersuchung fiel eine pigmentierte Läsion im Bereich der rechten Glutealregion auf, die deutliche klinische Atypien aufwies [ABB. 3A und B]. In der gezielten Anamnese berichtete die Patientin, dass sie erst kürzlich andernorts eine kostspielige KI-gestützte 3D-Ganzkörper-Hautkrebsvorsorge absolviert hatte. Bei diesem automatisierten Verfahren war die besagte gluteale Veränderung von der Software jedoch nicht als verdächtig detektiert oder für eine Folgeuntersuchung markiert worden. Wir rieten der Patientin dennoch zur Entfernung. Die histopathologische Aufarbeitung bestätigte die ärztliche Einschätzung: Es handelte sich um einen dysplastischen melanozytären Nävus vom Compound-Typ. Dieser klinische Fall zeigt, dass KI-gestützte Systeme mit computerbasiertem Mapping von Läsionen kritische Lücken aufweisen können.

Diskussion

Ein Grund für Fehlinterpretationen liegt möglicherweise in der mangelnden Transparenz der Datensätze, mit denen kommerzielle KI-Modelle trainiert werden (sogenannte «Black-Box-Modelle»)⁹. Für die Praxis bleibt oft unklar, ob seltene Melanom-Subtypen, atypische Lokalisationen oder unterschiedliche Hauttypen im Trainingskollektiv ausreichend repräsentiert waren¹⁰. Wenn ein Algorithmus vorwiegend mit klassischen Lehrbuch-Melanomen trainiert wurde, läuft er Gefahr, untypische oder morphologisch diskrete Verläufe, wie beispielsweise im zweiten Fall, fälschlicherweise als «unauffällig» einzustufen⁹.



[ABB. 2A] Auffällige Stelle an der rechten Flanke



[ABB. 2B] Auflichtmikroskopische Aufnahme der Läsion zwei Monate vor dem operativen Eingriff. Bei genauer Betrachtung der Morphologie zeigt sich bereits hier ein asymmetrisches Pigmentnetzwerk im unteren linken Quadranten.



[ABB. 2C] Präoperative Folgeaufnahme der identischen Stelle. Im direkten dermatoskopischen Vergleich zeigt der Befund eine weitgehende Konstanz ohne signifikante Dynamik.

Während im zweiten klinischen Fall ein malignes Melanom trotz unauffälliger Software-Klassifizierung histopathologisch gesichert wurde, offenbart Patientenfall 1 die akute Gefahr einer softwareseitigen Unterbewertung im klinischen Alltag. Unabhängig vom finalen histologischen Ergebnis führt die Diskrepanz zwischen ärztlicher Dringlichkeit und digitaler Entwarnung zu einem massiven Vertrauenskonflikt aufseiten der Patient:innen.

Fallbericht 3 illustriert eine ganz andere technologische Limitation: das Versagen bei der automatisierten Detektion im Rahmen des 3D-Ganzkörper-Mappings. Solche Systeme nutzen Algorithmen zur automatisierten Erkennung, Zählung und Beurteilung von Hautläsionen am gesamten Körper¹¹. Befindet sich eine Veränderung jedoch in einer anatomischen Region, die während des Scanvorgangs potenziell durch Unterwäsche verdeckt bleibt, stösst die rein kamerabasierte Automatisierung an ihre Grenzen.

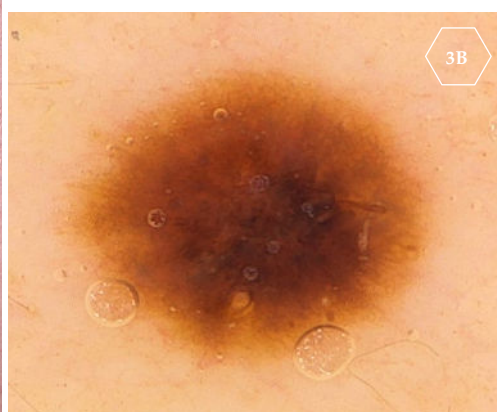
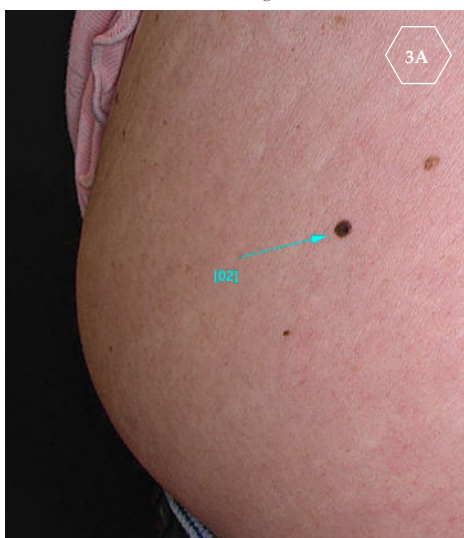
Trotz dieser evidenten Limitationen wäre es jedoch falsch, den Einsatz von KI in der Dermatologie künftig pauschal abzulehnen. Die Technologie sollte nicht als fehlerfreier Ersatz, sondern als Assistenzsystem verstanden werden. Wissenschaftliche Untersuchungen zeigen wiederholt, dass weder die KI allein noch das ärztliche Personal isoliert die besten diagnostischen Ergebnisse erzielen. Die höchste Sensitivität und Spezifität werden in der Symbiose aus fachärztlicher Expertise und technologischer

Unterstützung erreicht². Die KI kann dabei als sensitives «zweites Augenpaar» fungieren und subtile Veränderungen im Verlauf frühzeitig erkennen. Die abschliessende Beurteilung, unter Einbezug von Anamnese, Palpation und dem klinischen Gesamtbild (*ugly-duckling-sign*) sollte jedoch zwingend durch die Ärztin oder den Arzt erfolgen¹¹. ◊

Bibliografie

- ¹ Hartjen A, et al.: Künstliche Intelligenz in der Lasermedizin – Technologie mit enormem Potenzial. *Dermatologe* 2022; 70(2): 124–131.
- ² Laiouar-Pedari S, et al.: Prospective Evidence on Artificial Intelligence-Assisted Melanoma Diagnostics A Systematic Review and Meta-Analysis. *JAMA Dermatol* 2026; 162(5): 478–487.
- ³ Esteva A, et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115–118.
- ⁴ Brinker T, et al.: Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 2019; 119: 11–17.
- ⁵ Anriot J, et al.: Limits of Artificial Intelligence Models for Skin Cancer Diagnosis in Realistic Settings. *JAMA Dermatol* 2026; 162(7): 701–708.
- ⁶ Kather J: KI in der klinischen Routine. *Künstliche Intelligenz in der Medizin* 2025: 139–172.
- ⁷ Sokolov K, et al.: Dynamics of the Neural Network Accuracy in the Context of Modernization of the Algorithms of Skin Pathology Recognition. *Indian J Dermatol* 2022; 67(3): 312.
- ⁸ Haenssle H, et al.: Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol* 2020; 31: 137–143.
- ⁹ Grzybowski A, et al.: Challenges of artificial intelligence in medicine and dermatology. *Clin Dermatol* 2024; 42: 210–215.
- ¹⁰ Gottfrois P, et al.: PASSION for Dermatology: Bridging the Diversity Gap with Pigmented Skin Images from Sub-Saharan Africa. *MICCAI* 2024: 703–712.
- ¹¹ Baete F, et al.: 3D Total Body Photography as a Promising Innovation for Early Skin Cancer Detection: Scoping Review. *JMIR Dermatol* 2025; 8: e68510.

[ABB. 3A] Pigmentierte Läsion in der rechten Glutealregion; die Läsion hebt sich aufgrund ihrer Grösse und Farbe von den übrigen Stellen ab.



[ABB. 3B] Auf der dermatoskopischen Aufnahme ist das Pigmentnetzwerk durch die starke zentrale Hyperpigmentierung schlecht zu erkennen.