



Current Perspective

Prediction of melanoma evolution in melanocytic nevi via artificial intelligence: A call for prospective data



Wiebke Sondermann ^{a,1}, Jochen Sven Utikal ^{b,c,1}, Alexander H. Enk ^d, Dirk Schadendorf ^a, Joachim Klode ^a, Axel Hauschild ^e, Michael Weichenthal ^e, Lars E. French ^f, Carola Berking ^g, Bastian Schilling ^h, Sebastian Haferkamp ⁱ, Stefan Fröhling ^j, Christof von Kalle ^j, Titus J. Brinker ^{d,j,*}

^a Department of Dermatology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

^b Skin Cancer Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany

^c Department of Dermatology, Venereology and Allergology, University Medical Center Mannheim, Ruprecht-Karl University of Heidelberg, Mannheim, Germany

^d Department of Dermatology, University Hospital Heidelberg, Heidelberg, Germany

^e Department of Dermatology, University Hospital Kiel, University of Kiel, Kiel, Germany

^f Department of Dermatology, University Hospital Munich (LMU), Munich, Germany

^g Department of Dermatology, University Hospital Erlangen, University of Erlangen, Erlangen, Germany

^h Department of Dermatology, University Hospital Würzburg, University of Würzburg, Würzburg, Germany

ⁱ Department of Dermatology, University Hospital Regensburg, Regensburg, Germany

^j National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany

Received 26 May 2019; received in revised form 22 June 2019; accepted 4 July 2019

Available online 8 August 2019

KEYWORDS

Melanoma;
Skin cancer;
Artificial intelligence;
Deep learning;
Prediction

Abstract Recent research revealed the superiority of artificial intelligence over dermatologists to diagnose melanoma from images. However, 30–50% of all melanomas and more than half of those in young patients evolve from initially benign lesions. Despite its high relevance for melanoma screening, neither clinicians nor computers are yet able to reliably predict a nevus' oncologic transformation. The cause of this lies in the static nature of lesion presentation in the current standard of care, both for clinicians and algorithms. The status quo makes it difficult to train algorithms (and clinicians) to precisely assess the likelihood of a benign skin lesion to transform into melanoma. In addition, it inhibits the precision of current algorithms since 'evolution' image features may not be part of their decision. The current literature reveals

* Corresponding author: National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany. Fax: +496221 566967.

E-mail address: titus.brinker@nct-heidelberg.de (T.J. Brinker).

¹ These authors contributed equally to this work.

certain types of melanocytic nevi (i.e. ‘spitzoid’ or ‘dysplastic’ nevi) and criteria (i.e. visible vasculature) that, in general, appear to have a higher chance to transform into melanoma. However, owing to the cumulative nature of oncogenic mutations in melanoma, a more fine-grained early morphologic footprint is likely to be detectable by an algorithm. In this perspective article, the concept of melanoma prediction is further explored by the discussion of the evolution of melanoma, the concept for training of such a nevi classifier and the implications of early melanoma prediction for clinical practice. In conclusion, the authors believe that artificial intelligence trained on prospective image data could be transformative for skin cancer diagnostics by (a) predicting melanoma before it occurs (i.e. pre-in situ) and (b) further enhancing the accuracy of current melanoma classifiers. Necessary prospective images for this research are obtained via free mole-monitoring mobile apps.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Skin cancer is the most common cancer in white-skinned populations. Malignant melanoma, a skin cancer deriving from melanocytes, is accountable for most skin cancer-related deaths [1].

Early detection greatly improves the prognosis of patients with melanoma; however, despite the use of dermoscopy, dermatologists only rarely achieve sensitivities greater than 80% [2]. In 2017, Esteva *et al.* [30] were the first to report a deep-learning convolutional neural network (CNN) image classifier that performed on par with 21 board-certified dermatologists when identifying images with malignant lesions [3]. Subsequent artificial intelligence algorithms were able to demonstrate systematic outperformance of board-certified dermatologists [3–8] and most recently, our group demonstrated that the combination of human and artificial intelligence is superior to the individual results of dermatologists and CNNs [31].

While these developments show great potential to improve the standard of care, 30–50% of all melanomas and more than half of those in young patients evolve from initially benign lesions [9]. Despite its high relevance for patients undergoing skin checks, neither clinicians nor computers are yet able to reliably predict a melanocytic nevus’ oncologic transformation. Predicting melanoma in benign nevi would enable earlier detection and enhance prevention. In addition, all currently published algorithms were not trained with prospective data, making it impossible for them to learn image features that predict benign or malignant change. The cause of this lack of diagnostic ability lies in the static nature of lesion presentation in the current standard of care, both for clinicians and algorithms: Image sequences of the same lesion of the same patient are not regularly made at dermatology departments, many of which do not even conduct any photographic documentation of skin lesions (including university hospitals), and those that perform documentation do not take photographs in a sequential fashion or in standardised time frames.

For artificial intelligence algorithms, there are simply not enough labelled prospective training data available. The current literature reveals certain types of

melanocytic nevi (i.e. spitzoid/dysplastic nevi) and criteria (i.e. visible vasculature) that, in general, appear to have a higher chance to transform into melanoma, but the agreement between what is considered ‘spitzoid’ and ‘dysplastic’ between clinicians and histopathologists is low, and these criteria do not qualify for a fine-grained assessment. In this connection, two recent studies published in the *European Journal of Cancer* demonstrate that the precision of pathologists to diagnose melanoma may also be successfully enhanced by artificial intelligence algorithms [10,11].

1. Detecting early morphologic footprints of oncogenic mutations

Owing to the cumulative nature of oncogenic mutations in melanocytic nevi, a fine-grained early morphologic footprint should be detectable by an algorithm trained with prospective data: Differentiated melanocytes of the skin are described to have long cell life cycles and are unable to undergo mitosis [12]. Thus, melanocytic nevi may accumulate many mutations during their life cycle which are involved in various signalling cascades regulating proliferation (e.g. BRAF and NRAS), replication (TERT), cell cycle control (CDKN2A), metabolism (PTEN and KIT), apoptosis (TP53), etc. [13–17], and, conversely, physiologic pro-apoptotic responses to these (immune or intracellular responses) over long periods of time that likely leave morphologic footprints (i.e. small regression zones) which should be detectable in a high-resolution photo of the lesion.

2. Evidence on the importance of E for evolution in the ABCDE rule

In 1994, Nachbar *et al.* [18] first introduced the ABCD rule for dermoscopic melanoma detection which was based on four criteria which all demonstrated to increase the accuracy of dermoscopy: asymmetry, border, colour and diameter.

Since 1999, this method has been extended by a further criterion by Kittler *et al.* [19]: E for evolution. The authors incorporated information about morphologic changes of the lesion over time and demonstrated that the frequency of reported changes was significantly higher for melanomas than benign melanocytic nevi. Follow-up investigations confirmed that the clinical evolution is a crucial, if not one of the most important criteria when diagnosing skin lesions [20–22], and were shown to be of age-independent relevance [23]. However, owing to the high amount of patients an individual dermatologist sees, it is unlikely that he or she will remember the exact appearance of the lesion at the precluding skin check. A comparison of images of individual moles over time is already possible in clinical practice with applications offered by different software companies; however, they are not regularly used by clinicians, are not connected with prospective research efforts and come with considerable costs for the patients. In addition, most patients, especially those who are at risk for nevus-associated melanoma (i.e. with multiple melanocytic nevi) [9], will not necessarily remember if a lesion has changed or not. Furthermore, neither clinician nor current algorithms can fully assess whether a change itself was of malignant or benign character.

3. Enhancing the E of the ABCDE rule with prospective data sets

An algorithm trained on large data sets of prospective images of nevi remaining nevi and nevi transforming into melanoma over long periods of time would have the following predictive abilities:

- Detect a morphologic footprint which reveals the probability of future oncologic transformation of a nevus.
- If two images of the same nevus at two time points are available, assess the change of a nevus itself as rather benign or malignant.
- Precisely reveal the morphologic features predicting an oncogenic transformation in an understandable manner by the means of modern methods for interpreting and understanding deep neural networks [24].

All these aspects go beyond what is currently possible.

4. How to generate prospective data sets?

To yield the proposed functionality of machine learning, sequential images of many individual lesions from individual patients with follow-up images over

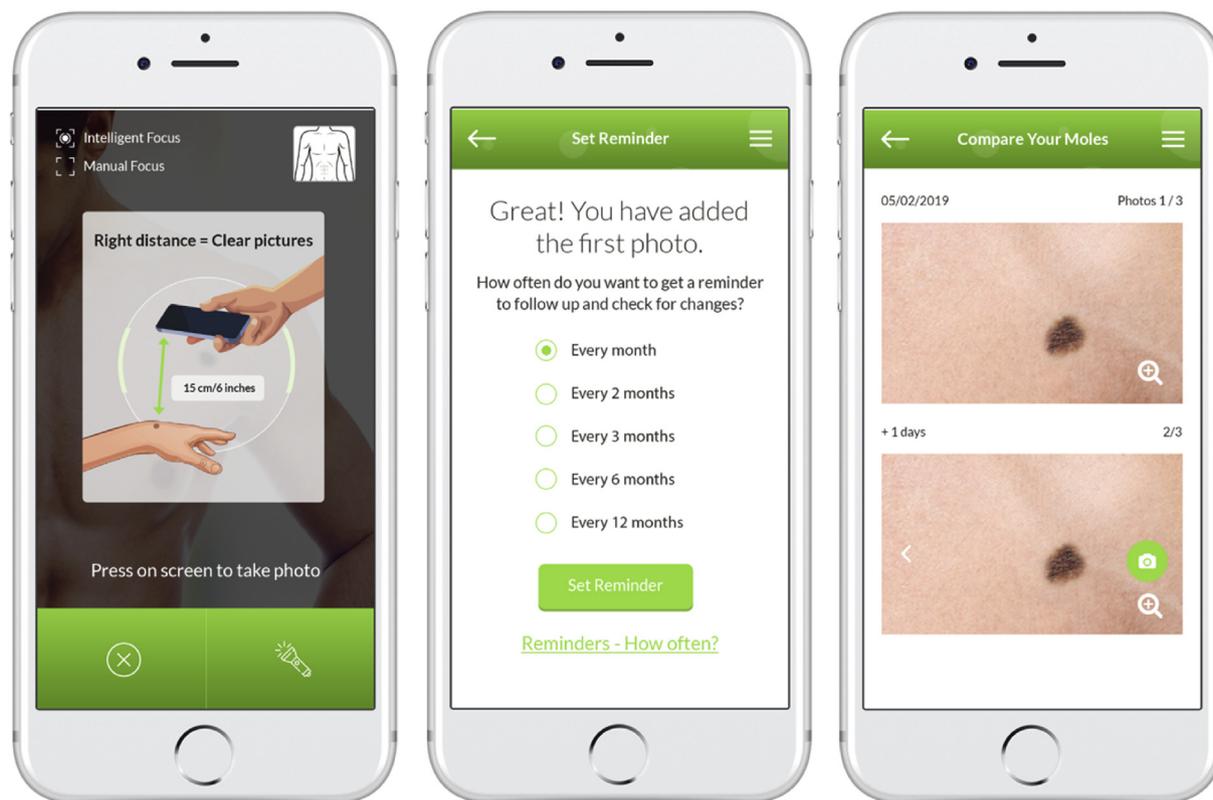


Fig. 1. Features of the plannedmole-monitoring app. Users can generate high-resolution images from specific distances (left), taken in specific time frames (middle), of the same mole(s) with a pin-pointed body location over a long period of time (right). All image sets are made exportable via an admin panel and are annotated by time frame, exact location and user ID. The app informs the patient that it is not a replacement for a visit at the physician but a help to self-monitor skin lesions which always should be shown to a dermatologist in case they change over time.

long time frames (i.e. months to years) are necessary. Mole-monitoring apps which acquire such prospective data by mapping moles over long periods of time via smartphone cameras have been around for years. However, the primary focus of these apps is to help smartphone users to monitor moles, but to date, no app that generates prospective image data specifically for research is available. Thus, the development of such a free research-based mole-monitoring app was initiated at the medical campus in Heidelberg together with colleagues from the Department of Dermatology at the University Hospital of Essen (Fig. 1; mock-ups of the apps prototype).

The app will enable to export prospective user-generated data sets for the training of an algorithm based on informed consent. User-generated sequential image sets will in turn be labelled with the precision by artificial intelligence algorithms trained on biopsy-verified images (and thereby exceeding dermatologists' precision) to be used for supervised learning of a deep neural network with the previously discussed abilities. An alternative would be to have the user self-label images after a biopsy was performed. Although this option will be offered within the app, it is deemed to be less reliable by the investigators because of a lack of compliance.

Of the vast majority of melanocytic nevi, only very few nevi will develop into melanoma [25], and this in turn is time dependent. Thus, the achievement of a large-enough data set (estimated to be at least 2000 user-generated nevi data sets with a combined follow-up period of approximately 1000 years for first assessments) with current data extraction strategies and based on past experiences of user acquisition by research-based apps launched by the investigators will take roughly three years [26–28]. Alternative strategies associated with substantially higher costs and efforts, however, would be the change of the current standard of care by enforcing digital follow-up images at standardised time intervals in dermatology departments.

5. The clinical value of melanoma prediction in benign nevi

While the overall prevalence of nevus-associated melanomas was described to be about 33% among all age groups [29], more than 50% of melanomas in young patients derive from initially benign nevi [9]. While malignant growth and metastasis can occur within weeks (especially in young patients), the costs for a comprehensive skin cancer screening in Europe are either not covered (i.e. not covered for young patients until the age of 35 years in Germany) or only covered in intervals of multiple years (i.e. every two years for people older than 35 years in Germany). Thus, predictive diagnostics, especially for young patients with numerous nevi, may reduce the number of melanoma-related deaths, specifically by enabling more focused

self-monitoring, early removal of nevi with a high probability of transformation but also by enhancing the precision of current melanoma classifiers already increasingly implemented in routine care.

Funding

This work received no funding.

Conflict of interest statement

None declared.

References

- [1] Schadendorf D, van Akkooi AC, Berking C, Griewank KG, Gutzmer R, Hauschild A, et al. Melanoma. *The Lancet* 2018; 392(10151):971–84.
- [2] Vestergaard M, Macaskill P, Holt P, Menzies S. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *Br J Dermatol* 2008;159(3):669–76.
- [3] Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur J Canc* 2019;113:47–54.
- [4] Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *European Journal of Cancer* 2019;111:148–54.
- [5] Brinker TJ, Hekler A, Hauschild A, Berking C, Schilling B, Enk AH, et al. Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. *Eur J Canc* 2019;111:30–7.
- [6] Brinker TJ, Hekler A, Enk A, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Canc* 2019; 119:11–7.
- [7] Maron R, Weichenthal M, Utikal J, Hekler A, Berking C, Hauschild A, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Canc* 2019;119:57–65.
- [8] Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* July 2019;20(7):938–47.
- [9] Pampena R, Pellacani G, Longo C. Nevus-associated melanoma: patient phenotype and potential biological implications. *J Invest Dermatol* 2018;138(8):1696–8.
- [10] Hekler A, Utikal J, Enk A, Berking C, Klode J, Schadendorf D, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Canc* 2019;113.
- [11] Hekler A, Utikal J, Enk A, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur J Canc* 2019 [in press].
- [12] Uong A, Zon LI. Melanocytes in development and cancer. *J Cell Physiol* 2010;222(1):38–41.
- [13] Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A landscape of driver mutations in melanoma. *Cell* 2012;150(2):251–63.

- [14] Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science* 2013;339(6122):957–9.
- [15] Read J, Wadt KA, Hayward NK. Melanoma genetics. *J Med Genet* 2016;53(1):1–14.
- [16] Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, et al. The genetic evolution of melanoma from precursor lesions. *N Engl J Med* 2015;373(20):1926–36.
- [17] Shain AH, Bastian BC. From melanocytes to melanomas. *Nat Rev Cancer* 2016;16(6):345.
- [18] Nachbar F, Stolz W, Merkle T, Cagnetta AB, Vogt T, Landthaler M, et al. The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. *J Am Acad Dermatol* 1994;30(4):551–9.
- [19] Kittler H, Seltenheim M, Dawid M, Pehamberger H, Wolff K, Binder M. Morphologic changes of pigmented skin lesions: a useful extension of the ABCD rule for dermatoscopy. *J Am Acad Dermatol* 1999;40(4):558–62.
- [20] Menzies S, Emery J, Staples M, Davies S, McAvoy B, Fletcher J, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. *Br J Dermatol* 2009;161(6):1270–7.
- [21] Haenssle HA, Krueger U, Vente C, Thoms K-M, Bertsch HP, Zutt M, et al. Results from an observational trial: digital epiluminescence microscopy follow-up of atypical nevi increases the sensitivity and the chance of success of conventional dermoscopy in detecting melanoma. *J Invest Dermatol* 2006;126(5):980–5.
- [22] Kittler H, Guitera P, Riedl E, Avramidis M, Teban L, Fiebiger M, et al. Identification of clinically featureless incipient melanoma using sequential dermoscopy imaging. *Arch Dermatol* 2006;142(9):1113–9.
- [23] Cordoro KM, Gupta D, Frieden IJ, McCalmont T, Kashani-Sabet M. Pediatric melanoma: results of a large cohort study and proposal for modified ABCD detection criteria for children. *J Am Acad Dermatol* 2013;68(6):913–25.
- [24] Montavon G, Samek W, Müller K-R. Methods for interpreting and understanding deep neural networks. *Digit Signal Process* 2018;73:1–15.
- [25] Damsky W, Bosenberg M. Melanocytic nevi and melanoma: unraveling a complex relationship. *Oncogene* 2017;36(42):5771.
- [26] Brinker TJ, Klode J, Esser S, Schadendorf D. Facial-aging app availability in waiting rooms as a potential opportunity for skin cancer prevention. *JAMA Dermatol* 2018;154(9):1085–6.
- [27] Brinker TJ, Seeger W, Buslaff F. Photoaging mobile apps in school-based tobacco prevention: the mirroring approach. *J Med Internet Res* 2016;18(6).
- [28] Brinker TJ, Brieske CM, Schaefer CM, Buslaff F, Gatzka M, Petri MP, et al. Photoaging mobile apps in school-based melanoma prevention: pilot study. *J Med Internet Res* 2017;19(9).
- [29] Haenssle HA, Mograby N, Ngassa A, Buhl T, Emmert S, Schön MP, et al. Association of patient risk factors and frequency of nevus-associated cutaneous melanomas. *JAMA Dermatol* 2016;152(3):291–8.
- [30] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115.
- [31] Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron R, et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Canc* 2019 [in press].