Dermatology

Nevi Article Series – Research Article

Dermatology
DOI: 10.1159/000517218

Received: November 10, 2020
Accepted: March 7, 2021
Published online: July 8, 2021

Reproducible Naevus Counts Using 3D Total Body Photography and Convolutional Neural Networks

Brigid Betz-Stablein[a, b] Brian D'Alessandro[c] Uyen Koh[b] Elsemieke Plasmeijer[a, d] Monika Janda[e] Scott W. Menzies[f, g] Rainer Hofmann-Wellenhof[h] Adele C. Green[a, i] H. Peter Soyer[b, j]

[a]QIMR Berghofer Medical Research Institute, Cancer and Population Studies, Brisbane, QLD, Australia; [b]The University of Queensland Diamantina Institute, The University of Queensland, Dermatology Research Centre, Brisbane, QLD, Australia; [c]Canfield Scientific Inc., Fairfield, NJ, USA; [d]Netherlands Cancer Institute, Dermatology Department, Amsterdam, The Netherlands; [e]Centre of Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia; [f]Sydney Medical School, The University of Sydney, Camperdown, NSW, Australia; [g]Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital, Camperdown, NSW, Australia; [h]Department of Dermatology, Medical University of Graz, Graz, Austria; [i]CRUK Manchester Institute and University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK; [j]Dermatology Department, Princess Alexandra Hospital, Brisbane, QLD, Australia

Keywords
Melanocytic naevi · Moles · Melanoma · Artificial intelligence · 3D total body imaging

Abstract
Background: The number of naevi on a person is the strongest risk factor for melanoma; however, naevus counting is highly variable due to lack of consistent methodology and lack of inter-rater agreement. Machine learning has been shown to be a valuable tool for image classification in dermatology. ***Objectives:*** To test whether automated, reproducible naevus counts are possible through the combination of convolutional neural networks (CNN) and three-dimensional (3D) total body imaging. ***Methods:*** Total body images from a study of naevi in the general population were used for the training (82 subjects, 57,742 lesions) and testing (10 subjects; 4,868 lesions) datasets for the development of a CNN. Lesions were labelled as naevi, or not ("non-naevi"), by a senior dermatologist as the gold standard. Performance of the CNN was assessed using sensitivity, specificity, and Cohen's kappa, and evaluated at the lesion level and person level. ***Results:*** Lesion-level analysis comparing the automated counts to the gold standard showed a sensitivity and specificity of 79% (76–83%) and 91% (90–92%), respectively, for lesions ≥2 mm, and 84% (75–91%) and 91% (88–94%) for lesions ≥5 mm. Cohen's kappa was 0.56 (0.53–0.59) indicating moderate agreement for naevi ≥2 mm, and substantial agreement (0.72, 0.63–0.80) for naevi ≥5 mm. For the 10 individuals in the test set, person-level agreement was assessed as categories with 70% agreement between the automated and gold standard counts. Agree-

This article is part of the Nevi Article Series.
Adele C. Green and H. Peter Soyer: joint senior authorship.

karger@karger.com
www.karger.com/drm

Correspondence to:
Brigid Betz-Stablein, b.betzstablein@uq.edu.au

ment was lower in subjects with numerous seborrhoeic keratoses. ***Conclusion:*** Automated naevus counts with reasonable agreement to those of an expert clinician are possible through the combination of 3D total body photography and CNNs. Such an algorithm may provide a faster, reproducible method over the traditional in person total body naevus counts.

Introduction

Medical imaging is becoming increasingly common as a tool for diagnosis and monitoring of disease. It is particularly suited to dermatology given the highly visual nature of skin disease. Because diagnosis often relies on subjective assessment by medical practitioners, the addition of sequential images can greatly assist in monitoring lesion evolution or progression. Melanoma is one such skin condition that could benefit greatly from accurate and objective skin monitoring, as the long-term prognosis is critically dependent on early detection [1].

The strongest known risk factor for melanoma is the number of melanocytic naevi [2, 3]. However, counting naevi is highly subjective, and there is no consensus on a standard methodology for counting. Reported naevus counts differ by body site studied, by size of naevi counted (>2 mm, >3 mm, >5 mm) and by who counts them (medical practitioners, trained researchers, self-report) [4, 5]. Consequently, objectively recorded counts from imaging could be beneficial in identifying people at high risk for melanoma; hence, a standardised, objective, and repeatable naevus counting algorithm is needed.

Traditionally, to monitor all naevi over a person's whole skin surface, 2D total-body photography has been employed. This is time and resource intensive and requires a photographer to take on average 24 photos of subjects in different poses [6]. The use of 2D total body imaging has been shown to reduce the number of biopsies taken, and increase the accuracy of diagnosis in people at high risk of melanoma [7]. The introduction of three-dimensional (3D) total body photography allows for the simultaneous capture of 92 images that are then constructed into a 3D avatar [8]. However, the validity and clinical application of 3D total body photography for lesion identification has not yet been demonstrated.

Given the large amount of data collected by 3D total body imaging, automatic methods for lesion detection and classification would greatly aid clinicians and researchers. Automated naevus counting has been attempted using the relatively uncomplicated images of pigmented lesions on children's backs [9]; however, the more complex task of distinguishing naevi amongst various types of pigmented lesions on adult photo-aged skin has not been undertaken. Recent research has shown that convolutional neural networks (CNNs) are in most cases able to classify dermoscopic images of pigmented lesions with improved overall accuracy when compared with a group of human experts [10], with the potential for human-computer collaboration to provide further improvements [11]. The aim of this study was to measure the reliability of a deep neural network to assist in the identification of naevi from 3D total body photography of the full skin lesion ecosystem, and provide a total-body naevus count to assist with melanoma risk stratification [2, 3].

Methods

A summary of the methods is shown in Figure 1.

Naevus-Counting Algorithm Development

Training and test datasets were randomly selected from the 3D avatars of subjects of the Mind Your Moles study [12], a population-based cohort study in Queensland, Australia. Subjects were imaged using a VECTRA®WB360 (Canfield Scientific Inc., Parsippany, NJ, USA), which simultaneously takes 92 images, combining them into a 3D avatar [8]. Demographic factors were collected using standard questionnaires with clinical characteristics collected by research assistants, as previously described [12].

For the training of CNNs, it is generally recommended that a minimum of 5,000 labelled images per category be provided [13]. The training set from the 3D avatars of 82 randomly selected subjects consisted of 57,742 automatically detected lesion images ≥2 mm in diameter (Fig. 1). The number of lesions per subject ranged from 59 to 4,125 (median of 539). The age of study population ranged from 23 to 69 (median of 55) years and 52 (62%) were male. All subjects had Fitzpatrick skin types I–IV. Seven subjects (9%) reported a personal history of melanoma. All lesions were labelled on-screen as either naevi (n = 5,106, 10%) or non-naevi (n = 52,636, 90%) by a senior dermatologist with extensive experience using the VECTRA system. An independent test set of images from an additional 10 subjects were labelled on-screen independently by three expert physicians with consensus calculated as agreement of ≥2 clinicians. Additionally, naevi were manually identified in clinic by the senior dermatologist using a dermatoscope – this was considered the gold standard for the test set. To match the in clinic counts with the automated 3D total body image counts, lesions under the underwear, on the scalp or soles of the feet were not counted. A minimum of 1 month was required between the on-screen and in-clinic labelling. For greater repeatability, only lesions ≥2 mm were considered for classification.

Convolutional Neural Network

A perspective-corrected tile image of each lesion was generated by reprojecting the original 2D views of the lesions using the calibrated 3D geometry information of the 3D avatar. This created a

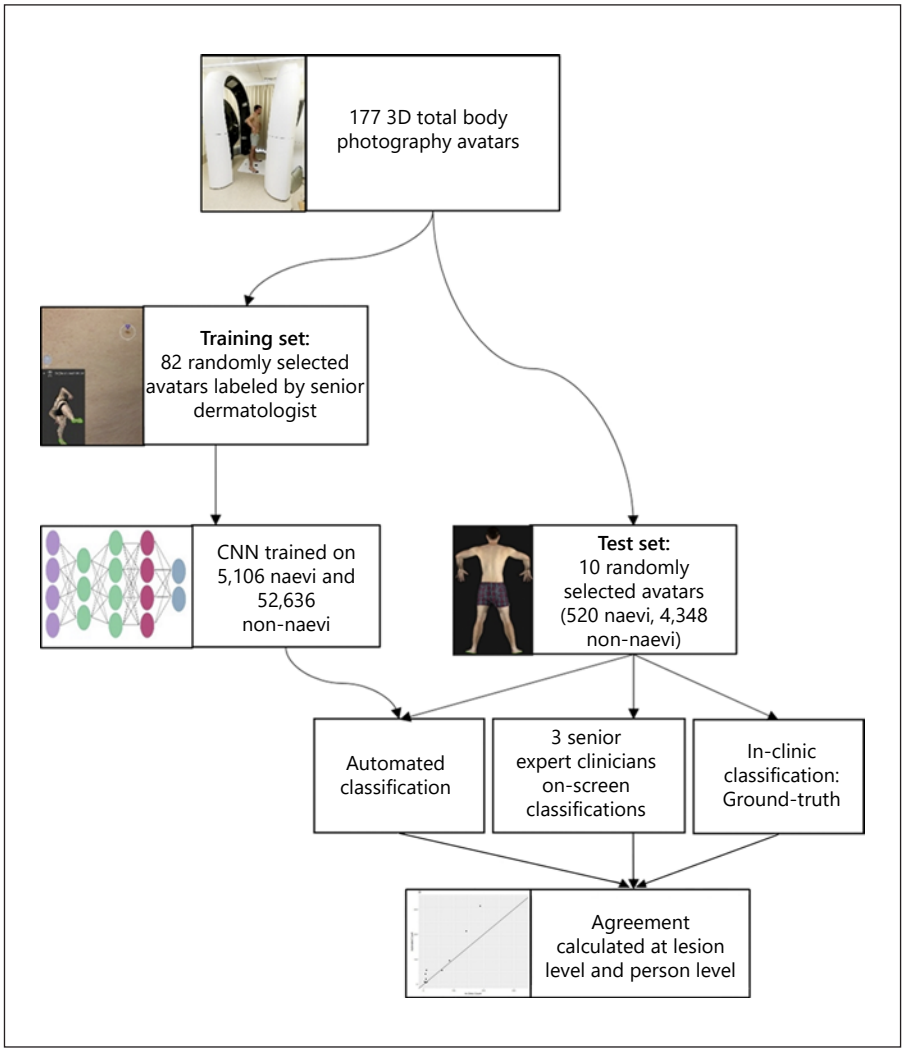

Fig. 1. Workflow from imaging to the automated naevus counts.

view of a lesion as if it were photographed perpendicularly to the skin surface. A two-class CNN classifier was then constructed which accepted a single tile image of a lesion, and outputs the probability that the lesion belonged to the "Naevus" and "Non-Naevus" classes. Three of the 82 training subjects were randomly selected to form a validation set to tune the hyperparameters of the CNN. The validation set therefore consisted of 1,023 lesion images (1.8%), including 248 naevi. Because of the class imbalance of the dataset, the images in the "Naevus" class were replicated 8 times to facilitate training. The full training data was further augmented through random rotations, flips, intensity, contrast, and colour adjustments to reduce overfitting. The CNN was constructed as three blocks of two convolutional layers, a pooling layer, and Rectified Linear Unit activation, followed by two fully connected layers. The network weights were initialised from a previously trained lesion analysis network. Training occurred using deep learning framework Caffe [14] on two Nvidia 1080Ti graphics processing units using the Root Mean Square Propagation optimiser over 500 epochs (iterations of the entire dataset) with a sigmoid decay learning rate scheduler.

Statistical Analysis

The performance of the algorithm and clinicians was compared at both the lesion level and the person level. The on-screen identifications of 3 expert clinicians across 10 test subjects were compared with the identifications from the automated VECTRA algorithm described above. Cohen's kappa [15], specificity, sensitivity, and overall agreement were calculated with the in-clinic classification used as the gold standard. At the person level, total counts ≥5 mm were compared using Bland-Altman Limits of Agreement. In addition, counts (≥2 mm) were categorised as "few" (<20), "some" (20–50), "many" (>50), and compared across counting methods.

Results

Test Dataset

The test set contained 10 subjects (5 male, 5 female) randomly selected from the total study cohort, with a me-

Table 1. Point estimates and 95% confidence intervals for automated naevus classification compared to the gold standard in-clinic naevus counts in the test set (N = 10; n = 4,868)

	Overall accuracy	Cohen's kappa	Sensitivity	Specificity	Positive predictive value	Negative predictive value
All lesions ≥2 mm (n = 4,868*)						
Clinician 1	0.90	0.45 (0.41–0.49)	0.50 (0.45–0.54)	0.94 (0.93–0.95)	0.51 (0.47–0.56)	0.94 (0.94–0.95)
3-clinician consensus	0.91	0.45 (0.41–0.49)	0.42 (0.38–0.46)	0.97 (0.96–0.97)	0.61 (0.56–0.66)	0.93 (0.93–0.94)
Automated	0.90	0.56 (0.53–0.59)	0.79 (0.76–0.83)	0.91 (0.90–0.92)	0.51 (0.47–0.54)	0.97 (0.97–0.98)
Lesions ≥2mm and <5mm (n = 4,480*)						
Clinician 1	0.90	0.44 (0.40–0.48)	0.51 (0.46–0.56)	0.94 (0.93–0.95)	0.48 (0.44–0.53)	0.95 (0.94–0.95)
3-clinician consensus	0.91	0.45 (0.40–0.49)	0.43 (0.38–0.47)	0.97(0.96–0.97)	0.58 (0.53–0.64)	0.94 (0.93–0.95)
Automated	0.89	0.54 (0.50–0.57)	0.78 (0.74–0.82)	0.91 (0.90–0.92)	0.48 (0.44–0.51)	0.97 (0.97–0.98)
Lesions ≥5 mm (n = 388*)						
Clinician 1	0.86	0.50 (0.39–0.61)	0.43 (0.33–0.55)	0.98 (0.95–0.99)	0.84 (0.69–0.93)	0.86 (0.82–0.90)
3-clinician consensus	0.86	0.47 (0.36–0.58)	0.40 (0.29–0.51)	0.98 (0.96–0.99)	0.85 (0.69–0.94)	0.86 (0.82–0.89)
Automated	0.90	0.72 (0.63–0.80)	0.84 (0.75–0.91)	0.91 (0.88–0.94)	0.73 (0.63–0.81)	0.96 (0.93–0.98)

* According to in-clinic gold standard naevus counts.

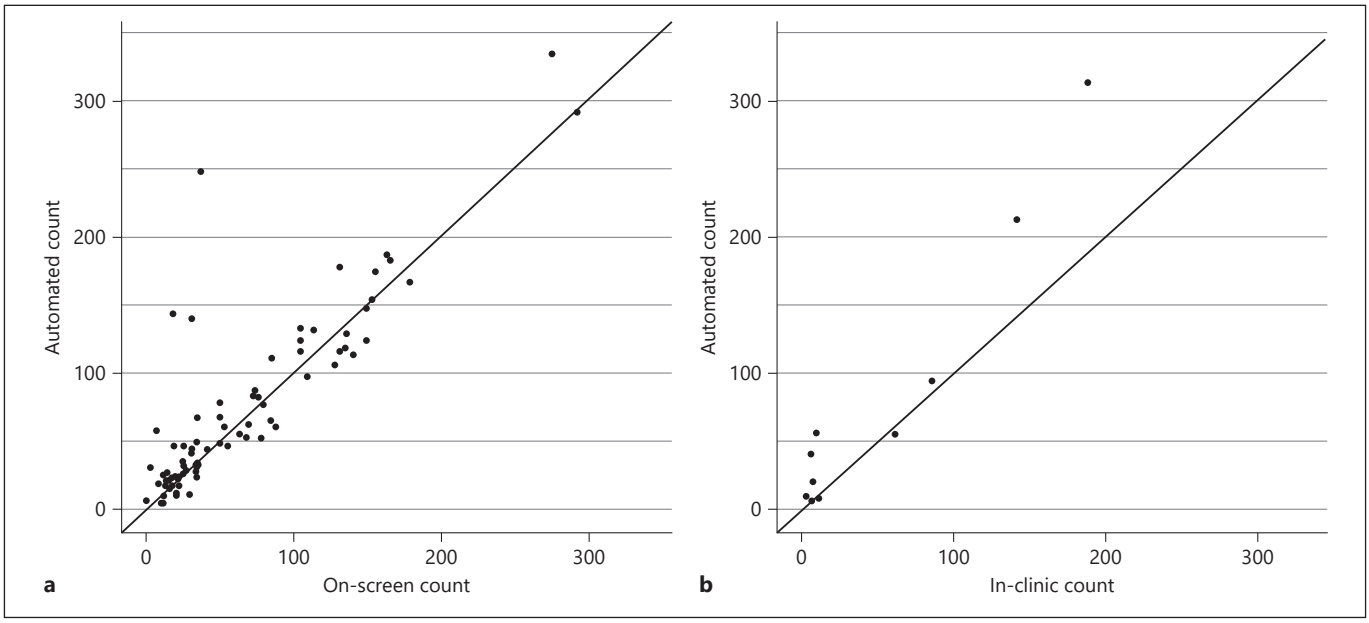

Fig. 2. Correlation between automated and on-screen counts for the training dataset (**a**) and between automated and in-clinic gold standard for the test dataset (**b**).

dian age of 57 (37–67) years. All subjects were of European descent, the majority with fair innate skin colour (n = 8, 80%) and blue/green eyes (n = 9, 90%). Fifty percent (n = 5) reported a history of keratinocyte cancer, and 3 (30%) reported a history of melanoma (all in situ) (suppl. Table 1; for all online suppl. material see www.karger.com/ doi/10.1159/000517218).

Lesion-Level Analysis in the Test Set

Of the 4,868 lesions ≥2 mm analysed across the 10 subjects, naevi identified in-clinic represented 11% (n = 520) of all lesions ≥2 mm, and 21% (n = 83) of the 388 lesions ≥5 mm, with the remainder being other benign lesions (e.g., solar lentigines, seborrhoeic keratoses, angiomas).

On-Screen Naevus Identification versus Gold Standard In-Clinic

Overall agreement for on-screen identification compared with gold standard in-clinic identification by clinician 1 (who performed the onscreen gold standard for the training set) was 90%. Cohen's kappa was 0.45 (0.41–0.49) indicating moderate agreement, with a sensitivity of 50% (45–54%) and a specificity of 94% (93–95%). A decrease in sensitivity and an increase in specificity were seen for both clinician 1 and the consensus of the clinicians (Table 1) for identifying naevi ≥5 mm. Comparable agreement was seen between the consensus of the three clinicians on-screen and the in-clinic counts (Table 1).

Automated versus Gold Standard In-Clinic Classification

The sensitivity and specificity of the automated CNN for lesions ≥2 mm was 79% (76–83%) and 91% (90–92%), respectively (Table 1). For lesions ≥5 mm, the automated algorithm showed a slightly higher sensitivity of 84% (75–91%), and similar specificity 91% (88–94%). Comparing the CNN with the in-clinic assessment by the dermatologist resulted in a Cohen's kappa of 0.56 (0.53–0.59), indicating moderate agreement for naevi ≥2 mm, and substantial agreement (0.72, 0.63–0.80) for naevi ≥5 mm.

Person-Level Analysis

Correlation between Automated and Clinician Total Body Naevus Counts

Excellent correlation was seen between AI counts and clinician on-screen tags for the training (87%, Fig. 2a), and between the AI counts and the gold standard test dataset (97%, Fig. 2b). However, the median difference between the automated and in-clinic counts was 11, with inter-quartile range of 4–19 for the training, and 6–44 for the test set. When naevus counts were converted to categories (few, some, many), the algorithm and in-clinic agreement was 70% (Table 2). In two (20%) of the mismatched cases, the expert dermatologist rated the naevus count in the few (<20), and the algorithm in the some (20–50) categories. In the third case, the dermatologist rated the count as few while the AI rated it as many (>50). The two subjects whose naevus counts were mismatched with the largest disparity between algorithm and dermatologist were both noted to have many seborrhoeic keratoses; this phenomenon was also seen in the training dataset (Fig. 3). When the two subjects with many seborrhoeic keratoses were ex-

Table 2. Total body naevus counts (≥2 mm) from the onscreen consensus between 3 clinicians, clinician 1 counts onscreen and in-clinic (gold standard), AI counts and self-report category

	Total body naevus count ≥2 mm								
	Raw counts				Naevus count category				
	self-report	3-clinician consensus	clinician 1 onscreen	clinician 1 in-clinic	automated counts	self-report	3-clinician consensus	clinician 1 onscreen	clinician 1 in-clinic	automatedd counts
a	0–20	112	141	142	213	Many	Many	Many	Many	Many
b	0–20	3	4	3	10	Few	Few	Few	Few	Few
c	20–50	139	203	188	313	Some	Many	Many	Many	Many
d	0–20	3	6	8	21	Few	Few	Few	Few	Some
e	50+	40	49	86	94	Many	Some	Some	Many	Many
f	0–20	4	7	11	8	Few	Few	Few	Few	Few
g*	50+	14	21	6	41	Many	Few	Some	Few	Some
h	20–50	13	22	6	6	Some	Few	Some	Few	Few
i*	20–50	7	24	9	56	Some	Few	Some	Few	Many
j	20–50	22	27	61	55	Some	Some	Some	Many	Many

Few, <20 naevi; some, 20–50 naevi; many, >50 naevi. * Participants **g** and **i** have numerous seborrhoeic keratoses.

Fig. 3. Images of back annotated with total body naevus counts by clinician in-clinic gold standard (Clin) and automated artificial intelligence counts (AI). Subjects in **g** and **i** had many seborrhoeic keratosis on their bodies, and AI varied proportionally the most.

cluded, the Bland-Altman limits of agreement showed 95% of the automated counts ≥5 mm were within ±5 of the in-clinic gold standard counts (79% of counts were within ±3).

Discussion

In this study, we developed and tested an algorithm for automated total body naevus counts from 3D total body photography based on 57,742 automatically detected skin lesions ≥2 mm in 82 study subjects. In addition to being the first algorithm applied to such data, it provides a time-saving, objective method to standardise naevus counting. Overall accuracy in the identification of individual naevi was high, with automated naevus counts showing reasonable agreement with the gold standard dermatologist counts in-clinic.

Measures of inter-rater reliability of total body naevus counts are rarely provided in the literature, likely due to the time-consuming nature of obtaining naevus counts on the set of same people by multiple times by multiple observers. Reproducibility of total naevus counts has been reported more frequently in children in whom naevi are easier to identify than in adults due to decreased levels of photo-damage and nearly complete lack of other benign lesions in youth that are common with aging such as solar lentigines and seborrhoeic keratoses [16]. In studies that have reported inter-rater reliability, it is often reported as a correlation which is generally high (e.g., naevi ≥2 mm: $r = 0.88$ [17], $r = 0.95$ [18]), but correlation does not accurately reflect agreement [19]. Due to low levels of agreement between naevus counts, counts are often categorised and agreement measured using Cohen's Kappa (κ). In adults, the inter-rater reliability for naevus counts ≥2 mm on the left arm between trained interviewers and dermatologist has been reported to be poor (four categories: 0, 1–4, 5–10, 11+, $\kappa = 0.19$) or moderate ($\kappa = 0.51$) between two dermatologists [20]. Inter-rater agreement between two dermatology residents for counts of naevi of any size was substantial (four categories: 1–10, 11–500, 51–100, >100; $\kappa = 0.66$). Similar results have been reported with self-classification of naevi (any size) compared with a dermatologist (five catego-

ries: 0, 1–9, 10–39, 40–100, >100, $\kappa = 0.14$), and for naevi ≥2 mm (5 categories: ≤5, 6–15, 16–30, 31–50, 51+, $\kappa = 0.19$ [21]), improving to $\kappa = 0.32$ with two categories (<50, ≥50) [22]. For total body naevus counts ≥5 mm, 79% of self-counts and physician counts agreed within ±3 counts [5]. This is similar to the agreement observed between the CNN-acquired automated counts and in-clinic gold standard.

The automated categorisation of the biological ecosystem of naevi and other benign skin lesions is of potential value for better understanding the development of melanoma and keratinocyte cancer. However, whether a specific lesion is a benign naevus or another type of benign lesion (e.g., solar lentigo or seborrhoeic keratosis) is of mainly academic interest; in a real-world clinical setting, a medical practitioner is generally only interested in whether a lesion is suspicious for malignancy or not. Therefore, the clinical utility of this algorithm – distinguishing between a benign naevus and another type of benign lesion ("non-naevus") – lies in its ability to provide instant and accurate naevus counts which can be used to objectively categorise a person's melanoma risk.

There is debate as to the best way to present naevus counts, and categorical counts have been suggested as the more generalisable [23]. Different cut-offs have been proposed [24] with no consensus being reached. Some studies have shown that an increase of even one naevus increases a person's melanoma risk [25]. The large variations in risk models and lack of reproducibility in naevus counting methods could thus lead to misclassification of melanoma risk [26].

While our approach is promising, there are limitations which should be improved in future work. This includes identifying naevi for the training set on-screen rather than in person; however, the gold standard for the test set was taken to be the in-clinic counts by a senior dermatologist. While the test set was small in terms of subjects ($n = 10$), the corresponding number of test lesions was large ($n = 4{,}868$). However, this meant we were unable to evaluate if the algorithm performs differently for different subject characteristics such as level of photodamage or sex. For example, in our test set the algorithm overcalled the counts in all 5 men, but this could just be due to chance. It is also necessary to further develop this algorithm for darker skin types not included in the study population. The algorithm performed poorly on people with many seborrhoeic keratoses, but such people can easily be identified in the clinic and flagged for manual counts. Alternatively, a classification algorithm trained to identify both naevi and seborrhoeic keratosis may overcome this problem. Another constraint is that, even for experienced dermatologists, the distinction between benign naevi and benign "non-naevi" between 2 and 5 mm in diameter on severely photo-damaged skin is sometimes impossible, and therefore even the in-clinic gold standard assessment itself may not be fully reliable.

In conclusion, a standardised, objective, and repeatable naevus counting methodology, particularly taking into consideration naevus size, is required to provide accurate assessment of naevus count as a risk factor for melanoma. 3D total body photography combined with an automated algorithm, such as the one presented in this study, could provide further insights into the biological ecosystem of naevi and other benign skin lesions and has potential value for better understanding the drivers of melanoma and keratinocyte cancer development.

Key Message

Automated, reproducible naevus counts can be calculated by applying machine learning to 3D total-body images.

Acknowledgments

The authors would like to acknowledge the study subjects, research assistants Kaitlin Nufer, Chantal Rutjes, Caitlin Horsham, Saira Sanjida and Montana O'Hara for their help with data management, and the Dermatology Research Centre Clinical Research Team for collecting the images.

Statement of Ethics

This study was approved by the Human Research Ethics Committee of Princess Alexandra Hospital (HREC/09/QPAH, February 7, 2009) and The University of Queensland (UQ2009001590, August 26, 2009) and was conducted in accordance with the Declaration of Helsinki. Subjects provided written consent after receiving a participant information and consent form.

Conflict of Interest Statement

H.P.S. is shareholder of e-derm consult GmbH and MoleMap by Dermatologists Pty Ltd. He provides teledermatological reports regularly for both companies. H.P.S. also consults for Canfield Scientific Inc., Revenio Research Oy and is an adviser of First Derm™.

Funding Sources

This work was supported by the National Health and Medical Research Council – Centre of Research Excellence scheme (Grant number: APP1099021). H.P.S. is also funded by the Medical Research Future Fund – Next Generation Clinical Researcher's Program Practitioner Fellowship (APP1137127).

Author Contributions

B.B.S. had full access to the data and takes responsibility for the integrity of the data and accuracy of the data analysis. B.D., H.P.S., M.J., U.K., E.P.: study concept and design. B.D. developed the AI algorithm. H.P.S., S.W.M., R.H.W., U.K.: acquisition of data. B.B.S., B.D.: analysis and interpretation of data. B.B.S., U.K., A.C.G., H.P.S.: drafting of manuscript. B.B.S.: statistical analysis. A.C.G., H.P.S.: study supervision.

References

1 Green AC, Olsen CM, Hunter DJ. Textbook of Cancer Epidemiology. Oxford: Oxford University Press; 2018
2 Usher-Smith JA, Emery J, Kassianos AP, Walter FM. Risk prediction models for melanoma: a systematic review. Cancer Epidemiol Biomarkers Prev. 2014;23(8):1450–63.
3 Olsen CM, Pandeya N, Thompson BS, Dusingize JC, Webb PM, Green AC, et al. Risk Stratification for Melanoma: Models Derived and Validated in a Purpose-Designed Prospective Cohort. J Natl Cancer Inst. 2018;110:1075–83.
4 Gallagher RP, McLean DI. The epidemiology of acquired melanocytic nevi. A brief review. Dermatol Clin. 1995;13(3):595–603.
5 Lawson DD, Moore DH 2nd, Schneider JS, Sagebiel RW. Nevus counting as a risk factor for melanoma: comparison of self-count with count by physician. J Am Acad Dermatol. 1994;31(3 Pt 1):438–44.
6 Dengel LT, Petroni GR, Judge J, Chen D, Acton ST, Schroen AT, et al. Total body photography for skin cancer screening. Int J Dermatol. 2015;54(11):1250–4.
7 Truong A, Strazzulla L, March J, Boucher KM, Nelson KC, Kim CC, et al. Reduction in nevus biopsies in patients monitored by total body photography. J Am Acad Dermatol. 2016;75(1):135–e5.
8 Rayner JE, Laino AM, Nufer KL, Adams L, Raphael AP, Menzies SW, et al. Clinical Perspective of 3D Total Body Photography for Early Detection and Screening of Melanoma. Front Med (Lausanne). 2018;5:152.
9 Lee TK, Atkins MS, King MA, Lau S, McLean DI. Counting moles automatically from back images. IEEE Trans Biomed Eng. 2005;52(11):1966–9.
10 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. 2019;20(7):938–47.
11 Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229–34.
12 Koh U, Janda M, Aitken JF, Duffy DL, Menzies S, Sturm RA, et al. Mind your Moles' study: protocol of a prospective cohort study of melanocytic naevi. BMJ Open. 2018;8(9):e025857.
13 Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Cambridge: MIT Press; 2016.
14 Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. et al. Caffe: Convolutional Architecture for Fast Feature Embedding. Proceedings of the 22nd ACM International Conference on Multimedia; 2014.p. 675–678.
15 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
16 Yeatman JM, Kilkenny M, Marks R. The prevalence of seborrhoeic keratoses in an Australian population: does exposure to sunlight play a part in their frequency? Br J Dermatol. 1997;137(3):411–4.
17 Grulich AE, Bataille V, Swerdlow AJ, Newton-Bishop JA, Cuzick J, Hersey P, et al. Naevi and pigmentary characteristics as risk factors for melanoma in a high-risk population: a case-control study in New South Wales, Australia. Int J Cancer. 1996;67(4):485–91.
18 Bataille V, Snieder H, MacGregor AJ, Sasieni P, Spector TD. Genetics of risk factors for melanoma: an adult twin study of nevi and freckles. J Natl Cancer Inst. 2000;92(6):457–63.
19 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10.
20 Byles JE, Hennrikus D, Sanson-Fisher R, Hersey P. Reliability of naevus counts in identifying individuals at high risk of malignant melanoma. Br J Dermatol. 1994;130(1):51–6.
21 Buettner PG, Garbe C. Agreement between self-assessment of melanocytic nevi by patients and dermatologic examination. Am J Epidemiol. 2000;151(1):72–7.
22 Carli P, De Giorgi V, Nardini P, Mannone F, Palli D, Giannotti B. Melanoma detection rate and concordance between self-skin examination and clinical evaluation in patients attending a pigmented lesion clinic in Italy. Br J Dermatol. 2002;146(2):261–6.
23 Morze CJ, Olsen CM, Perry SL, Jackman LM, Ranieri BA, O'Brien SM, et al. Good test-retest reproducibility for an instrument to capture self-reported melanoma risk factors. J Clin Epidemiol. 2012;65(12):1329–36.
24 Ribero S, Zugna D, Osella-Abate S, Glass D, Nathan P, Spector T, et al. Prediction of high naevus count in a healthy U.K. population to estimate melanoma risk. Br J Dermatol. 2016;174(2):312–8.
25 Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. Eur J Cancer. 2005;41(1):28–44.
26 Betz-Stablein B, Koh U, Plasmeijer EI, Janda M, Aitken JF, Soyer HP, et al. Self-reported naevus density may lead to misclassification of melanoma risk. Br J Dermatol. 2020;182(6):1488–90.