Assessment of the change in accuracy of an artificial intelligence algorithm for the detection of skin cancer in camera images following diversification and training

Michael Phillips (michael.phillips@perkins.uwa.edu.au)
Harry Perkins Institute of Medical Research https://orcid.org/0000-0002-0252-9085
Jack Greenhalgh
Skin Analytics Limited, London

Technical advance

Keywords: Artificial intelligence, Machine learning, Image recognition, Skin cancer

Posted Date: September 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-78143/v1

Abstract

Background The US FDA recently stated in its Proposed Regulatory Framework for software as a medical device (SaMD) that "One of the greatest benefits of AI/ML in software resides in its ability to learn from real-world use and experience, and its capability to improve its performance." This study follows two previous publications which addressed the accuracy of a machine learning algorithm for the detection of malignant melanoma.

The aim of this study was to quantify the change in the accuracy following modifications to the algorithm (DERM) for the detection of non-melanoma skin cancers and potential precursors of skin cancer. A secondary aim was to assess any improvement in accuracy associated with continued training of the algorithm.

Methods A total of 16,550 images of skin lesions with histopathology based assessment were available for assessment. The primary indicator of diagnostic accuracy was the area under the ROC curve with 95% confidence intervals. Sensitivity and specificity at the most efficient cut-point was also estimated together with the numbers of false negative and false positive results.

Results The inclusion of squamous cell cancer, basal cell cancer and intra-epidermal carcinoma in addition to melanoma results in an improvement in the scope of the algorithm. For the most recent version of the algorithm all skin cancers show an area under the ROC curve greater than 95%. For melanoma sensitivity=91% and specificity = 89%; for all non-melanoma skin cancers sensitivity=97% and specificity=94%. Continued training of the algorithm results in a statistically significant ($p<0.01$) improvement in accuracy which diminishes as the ROC area approaches 100%.

Conclusions The results indicate that as the algorithm is used in clinical practice it will become more accurate with continued training but the rate of improvement will diminish as the ROC area approaches 100%.

A smartphone or other camera fitted with a dermoscopic lens and with internet access to the algorithm can provide an accurate additional assessment of a suspected skin cancer lesion or precursor for primary care physicians and dermatologists.

Introduction

The skin is one of the largest of the body organs. It is frequently a target for cancer because of exposure to solar ultra-violet radiation, with an estimate of more than one million non-melanoma skin cancers and nearly 300,000 melanomas worldwide in 2018.(1) The data for non-melanoma skin cancer is not routinely reported in many areas because of poor reporting which results in an underestimate of the health service costs associated with these relatively low mortality cancers. Basal cell carcinoma (BCC) is the most common of the non-melanoma skin cancers with one estimate of 80% BCC and 20% squamous cell carcinoma (SCC) in the United States(2). Variation in the incidence of both skin cancer types between countries is marked and is related to both ethnic differences and climatic factors which influence solar UV radiation as well as the damage done to the ozone layer by chlorofluorocarbon pollution. For these reasons Australia and New Zealand have the highest skin cancer incidence (age-standardised per 100,000) with melanoma incidence at 33.6 and 33.3 per 100,000 and non-melanoma incidence at 147.5 and 138.4 per 100,000 for each of those countries. For the USA the rates are 12.7 (melanoma) and 55.8 (non-melanoma) per 100,000. (3) There is a marked sex differential for non-melanoma incidence in all three countries with men more than twice as likely to develop the disease, possibly as a consequence of occupational exposure because of working outdoors exposed to sunlight.(4)

Since the publication of the US Surgeon General's Call to Action to Prevent Skin Cancer in 2014(2) a high priority has been assigned to skin cancer prevention in the United States with an emphasis upon melanoma. The most recent Skin Cancer Prevention Progress Report (5) indicates that there has been some progress in safe sun exposure practices but there is still a lack of primary prevention with one in three adults and more than half of high school students reporting sunburn each year(6). Even if there was a rapid improvement in sun-protection behaviour and facilities there is a latent reservoir of damaged skin from past sun exposure which will generate many skin cancers into the future.

As a consequence early detection remains a potential secondary prevention intervention but the current recommendation on skin cancer screening from the U.S. Preventive Services Task Force assessment (2016) concludes "that the current evidence is

insufficient and that the balance of benefit and harms of visual skin examination by a clinician to screen for skin cancer in asymptomatic adults cannot be determined."(7) The report indicates that one of the reasons for this conclusion is that detection of melanoma in primary care is not sufficiently accurate to support a population-based screening program.

Another issue which has created problems for an assessment of the value of screening for melanoma has been discussed by Weyers and that concerns overdiagnosis. As he demonstrates there is disagreement between dermatology clinicians and epidemiologists on what constitutes overdiagnosis. This has resulted in two quite different perspectives on melanoma detection – particularly in the context of small lesions (less than 6 mm.).(8)

Skin Analytics Ltd. has been developing an artificial intelligence based algorithm 'Deep Ensemble for the Recognition of Malignancy' (DERM) for the classification of skin cancer lesions based upon images captured by readily available cameras. The initial phase of the project was devoted to the detection of malignant melanomas and the results of the evaluation of these studies have been published.(9, 10) The studies showed that the DERM algorithm was as accurate as specialist dermatologists in detecting melanoma. The aim of the most recent development was to increase the scope of the algorithm to include a much wider range of skin lesions including non-melanoma skin cancers (SCC, BCC and intra-epidermal carcinomas) and lesions that may be precursors or be mistaken for skin cancers. (The list of the lesions can be seen in Table 1.)

Even though the US Food and Drug Administration states that "One of the greatest benefits of AI/ML in software resides in its ability to learn from real-world use and experience, and its capability to improve its performance" the statement is largely based upon non-medical commercial software apps and there is relatively little academic literature to describe the improvement that occurs with use and continuing training.

This paper presents the results of an up-dated assessment of the accuracy of the DERM algorithm following diversification. It also assesses the influence of continued retraining of the algorithm with newly acquired images.

Methods

A total of 16,550 images of skin lesions with histopathology based assessment were obtained from a variety of sources which included: Skin Analytics Tele-Dermatology Service (11), Skin Analytics Melanoma Prospective Study (10), PH2 Data Set (12), MoleMap Data Set (13), The ISIC Archive (14), The Interactive Atlas of Dermoscopy (15), The Kittler Data Set (16), and DermNet skin disease atlas (17). An additional 434 images of intact healthy skin were also used during the assessment and these were assumed to be negative with respect to histopathology for all lesion types.

DERM was designed and developed using deep learning techniques, specifically convolutional neural networks (CNNs) that can identify and assess features of skin lesions which are associated with each image type. Deep learning identifies features of a lesion directly from the data and contrasts the features that are associated with a positive compared to a negative histopathology assessment. Cross-validation was used to assess the performance of the algorithm; this approach allows every image to be assessed once, while ensuring the same image does not appear in the training and test dataset. Cross-validation is performed by splitting the dataset into 10 randomly sampled 'folds' (datasets). The algorithm is tested against each fold, with the remainder used for training. The results for each fold are then averaged so that the overall performance can be analysed. This method also avoids problems which result in overfitting.(18)

As images of lesions with histopathology assessment became available the algorithm was retrained on three subsequent occasions. This provided an opportunity to assess the improvement of the accuracy with continuing retraining. There were also changes in the algorithm which broadened the scope to include non-melanoma skin cancers and to specify potential precursors and specific types of benign lesions so that the algorithm cycles through the series of lesions until it identifies the specific type in a pre-determined sequence, which is illustrated in Fig. 1.

For this investigation we also assessed the accuracy of an algorithm based upon Google's Inception-V4 CNN Architecture (19) which allows a comparison between our trained versions of the algorithm and an existing pre-trained naive approach.

Receiver Operator Characteristic curves (ROC) with bootstrapped estimation (1,000 repetitions) were used to examine the overall diagnostic accuracy of the algorithm for each cancer type and the precursors (20). Area under the ROC (AUROC) was regarded as the most appropriate overall indicator of accuracy. Sensitivity, specificity and other diagnostic indicators were estimated for each lesion type at the most efficient decision threshold, where the threshold was determined as the point on the continuum which provided the closest balance between sensitivity and specificity. It should be noted that this approach assumes that false positives and false negatives have equal 'value' and this is unlikely to be valid in a clinical context but it is appropriate for the assessment of accuracy in the context of this study. The analysis was informative but it does not accommodate the way in which the algorithm might be used in practice as a 'virtual dermatologist'. In this clinical context any assessment that produced a positive result for melanoma would be referred for biopsy and histopathology assessment. If the initial result was negative for melanoma then the algorithm would assess for SCC and continue sequentially through the severity ordered lesions shown in Fig. 1. This approach allows each stage of assessment to occur and reflects potential severity for each outcome. We simulated this process for a more realistic mimic of the algorithm application. And used this severity ordered sequential assessment to assess the effective sensitivity and specificity for the skin cancers and precursor lesions.

A p-value less than 0.05 was regarded as statistically significant. All analysis was conducted using Stata Version 16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.).

Results

The distribution of the 16,550 lesions is shown in Table 1. The most frequent were benign lesions of various types (57.8%) followed by melanoma (14.5%) and BCC (7.50%). The total skin cancer lesions was 5,042 (30.0%).

Table 1
Frequency of assessed skin lesions

Category	Lesion type	Frequency (%)
Skin cancer	Melanoma	2,404 (14.5)
	Squamous Cell Carcinoma (SCC)	826 (4.99)
	Basal Cell Carcinoma (BCC)	1,242 (7.50)
	Intra-epidermal carcinoma	570 (3.44)
Potential precursor	Actinic Keratosis	1,193 (7.21)
	Dysplastic Nevus	757 (4.57)
Benign skin lesion	Benign Vascular Lesion	371 (2.24)
	Seborrheic Keratosis	1,500 (9.06)
	Dermatofibroma	260 (1.57)
	Lentigo	691 (4.18)
	Benign Melanocytic Nevus	1,529 (9.24)
	Other Benign	5,207 (31.5)
Total		16,550 (100)

Table 2 shows the accuracy of the assessment of the algorithm for each of the lesion types. The least accurate assessment was for melanoma (AUROC = 0.952) and the most accurate was for dermatofibroma (AUROC = 0.994). All of the AUROC estimates were greater than 95%.

Table 2
Accuracy of DERM for identification of each specific skin lesion

Lesion type	ROC area	LCL	UCL	Sensitivity	Specificity	TN (%)	TP (%)	FN (%)	FP (%)
Melanoma	0.952	0.948	0.956	88.6	88.3	12,868 (75.8)	2,129 (12.5)	275 (1.62)	1,712 (10.1)
SCC	0.982	0.978	0.985	94.7	92.7	14,974 (88.2)	782 (4.60)	44 (0.26)	1,184 (6.97)
BCC	0.987	0.984	0.989	94.0	94.9	14,937 (88.0)	1,168 (6.88)	74 (0.44)	805 (4.74)
Intra-epidermal carcinoma	0.975	0.970	0.978	92.3	100	14,926 (96.3)	526 (3.39)	44 (0.28)	0 (0.00)
Actinic keratosis	0.981	0.978	0.983	94.1	93.3	14,727 (86.7)	1,123 (6.61)	70 (0.41)	1,064 (6.26)
Dysplastic nevus	0.966	0.960	0.971	32.0	99.5	14,764 (86.9)	687 (4.04)	70 (0.41)	1,463 (8.61)
Benign vascular lesion	0.994	0.991	0.996	40.7	99.9	16,094 (94.8)	356 (2.10)	15 (0.09)	519 (3.06)
Seborrheic keratosis	0.975	0.971	0.978	90.4	92.6	14,338 (84.4)	1,356 (7.98)	144 (0.85)	1,146 (6.75)
Dermatofibroma	0.990	0.985	0.994	95.8	95.2	15,922 (93.8)	249 (1.47)	11 (0.06)	802 (4.72)
Lentigo	0.963	0.956	0.968	89.7	89.7	14,622 (86.1)	620 (3.65)	71 (0.42)	1,671 (9.84)
Benign melanocytic lesion	0.993	0.992	0.994	96.5	95.8	14,811 (87.2)	1,476 (8.69)	53 (0.31)	644 (3.79)
Other benign lesion	0.957	0.955	0.960	77.5	94.5	4,599 (27.1)	10,443 (61.5)	608 (3.58)	1,334 (7.85)

TN: True negative, TP: True positive, FN: False negative, FP: False positive

The severity ordered sequential assessment is shown in Table 3. It can be seen that all lesions are assessed for melanoma and those that are negative for melanoma are assessed for SCC *et seq*., the number of lesions assessed at each step can be seen from the column 'N'. It is clear that the deletion of the lesions that were positive for melanoma has significantly improved the accuracy of DERM for BCC and intra-epidermal carcinoma but not for SCC.

Table 3
Accuracy of DERM for identification of each skin lesion with sequential deletion of positive lesions previously detected as positive

Lesion	N	ROC area	LCL	UCL	Sensitivity	Specificity	TN (%)	TP (%)	FN (%)	FP (%)
Melanoma	16,984	0.952	0.948	0.956	88.6	88.3	12,868 (75.8)	2,129 (12.5)	275 (1.62)	1,712(10.1)
SCC	13,143	0.981	0.977	0.984	95.8	90.9	11,221 (85.4)	769 (5.85)	34 (0.26)	1,119 (8.5)
BCC	11,255	0.995	0.994	0.997	97.6	97.4	10,077 (89.5)	886 (7.87)	22 (0.20)	270 (2.40)
Intra-epidermal carcinoma	9,865	0.935	0.896	0.967	93.5	100	9,680 (98.1)	173 (1.75)	12 (0.12)	0 (0)
Actinic keratosis	9,926	0.992	0.988	0.994	96.9	97.3	8,952 (90.2)	671 (6.76)	22 (0.22)	281 (2.83)
Dysplastic nevus	8,974	0.970	0.963	0.976	39.1	99.4	7,725 (86.1)	469 (5.23)	45 (0.50)	735 (8.19)
Benign vascular lesion	8,718	0.996	0.993	0.997	60.4	99.9	8,110 (93.0)	328 (3.76)	6 (0.07)	274 (3.14)
Seborrheic keratosis	8,116	0.981	0.978	0.984	96.2	91.1	6,195 (76.3)	1,181 (14.6)	59 (0.73)	681 (8.39)
Dermatofibroma	6,254	0.996	0.994	0.997	97.8	95.9	5,820 (93.1)	177 (2.83)	4 (0.06)	253 (4.05)
Lentigo	5,824	0.984	0.977	0.989	92.2	97.6	5,320 (91.5)	228 (3.91)	22 (0.38)	247 (4.24)
Benign melanocytic lesion	5,349	0.994	0.993	0.996	99.8	93.9	3,912 (73.1)	1,178 (22.0)	3 (0.06)	256 (4.79)
Other benign lesion	3,915	0.971	0.964	0.977	94.5	98.5	448 (12.1)	3,082 (82.9)	181 (4.87)	7 (0.19)

TN: True negative, TP: True positive, FN: False negative, FP: False positive

Table 4 summarizes the overall performance with respect to false negatives and positives from the melanoma assessment. Twenty seven of the false negative melanoma results were assessed as true SCC (9.82%), 31 (11.3%) were assessed as true BCC and none were assessed as true Intra-epidermal carcinoma so that 21% of the total 275 false negative findings would be referred for excision and biopsy even though their true melanoma was not detected. A further 101 were assessed by the algorithm as actinic keratosis (n = 17, 6.03%) or dysplastic nevus (n = 84, 30.6%). Most of these would be removed using surgical or pharmaceutical treatments otherwise they would be monitored according to current clinical practice.(21, 22) Overall 159 of the 275 histopathology positive melanomas missed by DERM would be managed in a clinically appropriate manner (0.96% of the total images; 3.2% of all cancer images and 57.8% of the histopathology positive melanomas missed by DERM).

Table 4
Assessment of incorrect results from melanoma lesion review

Biopsy confirmed status	Total cases (%)	Melanoma FN	Melanoma FP
		Positive for alternative lesion (%)	Positive for alternative lesion (%)
SCC	826 (4.99)	27 (9.82)	49 (2.86)
BCC	1,242 (7.50)	31 (11.3)	96 (5.61)
Intra-epidermal carcinoma	570 (3.44)	0 (0.00)	25 (1.51)
Actinic keratosis	1,193 (7.21)	17 (6.20)	50 (2.92)
Dysplastic nevus	757 (4.57)	84 (30.6	532 (31.1)
Benign vascular lesion	371 (2.24)	14 (5.09	58 (3.39)
Seborrheic keratosis	1,500 (9.06)	43 (15.6	244 (14.3)
Dermatofibroma	260 (1.57)	24 (8.73	120 (7.01)
Lentigo	691 (4.18)	45 (16.4	510 (29.8)
Benign melanocytic lesion	1,529 (9.24)	19 (6.91	17 (0.99)
Other benign lesions	5,207 (31.5)	0 (0.00)	928 (54.2)
All lesions	16,550 (100)	304 (100)	2,629 (100)
Note: 275 lesions were FN and 1,712 were FP from melanoma review			

Table 4 also shows the true status of the false positives incorrectly identified as melanoma by DERM. These errors do not carry the same risk for patients as the false negatives but they may represent a cost to the health system or to patients because of inappropriate or invasive subsequent interventions. Of the 1,712 false positives 170 (9.93%) were SCC, BCC, Intra-epidermal carcinoma and 582 (34.0%) were AK or DN so that the subsequent management would have been appropriate despite the incorrect melanoma result.

Based only upon the melanoma assessment, DERM showed: sensitivity = 88.6 (95% CI: 87.2–89.8); specificity = 88.3 (87.7–88.8). Adjusted for a subsequent positive cancer result (SCC, BCC, Intra-epidermal carcinoma) DERM showed: sensitivity = 91.1 (89.8–92.1); specificity = 89.4 (88.6–89.9).

Table 5 shows the contrast between the Inception V4 algorithm and the initially trained version of the DERM algorithm. A version of the Inception V4 CNN which had been pre-trained to perform large-scale image recognition was retrained using the same data set which was used to train the latest version of DERM. It is clear that the DERM CNN vastly outperforms the Inception V4 CNN at the task of identifying skin lesions from dermoscopic images.

Table 5
Comparison of the area under the ROC curve between the Inception V4 and DERM

Lesion type	Algorithm type		X^2_1	*p*
	Inception V4	DERM		
Melanoma	0.8403	0.9461	869	< 0.0001
Squamous Cell Carcinoma (SCC)	0.9435	0.9805	205	< 0.0001
Basal Cell Carcinoma (BCC)	0.9149	0.9855	501	< 0.0001
Intra-epidermal carcinoma	0.9064	0.9721	243	< 0.0001
Actinic Keratosis	0.9360	0.9668	138	< 0.0001
Dysplastic Nevus	0.8552	0.9606	276	< 0.0001

Table 6 shows the AUROC for four development versions of the DERM algorithm in sequential order. Differences in the algorithm between development versions include improvements to the training methodology, changes to the neural network architecture, and the inclusion of additional training data. All versions of the DERM algorithm were assessed using the same data set, although the older versions used less training data. There was a statistically significant improvement in the area under the ROC curve over time for melanoma and for actinic keratosis and dysplastic nevus but not for BCC, SCC and Intra-epidermal carcinoma. The latter three lesions begin with very high levels of accuracy so this lack of improvement in accuracy with further training may be a consequence of a ceiling effect.

Table 6
Change in the area under the ROC curve with further training

Lesion type	Development Version of DERM				X^2_3	*p*
	1	2	3	4		
Melanoma	0.9461	0.9473	0.9478	0.9517	39.5	< 0.0001
Squamous Cell Carcinoma (SCC)	0.9805	0.9805	0.9816	0.9819	6.99	0.072
Basal Cell Carcinoma (BCC)	0.9855	0.9853	0.9855	0.9867	7.66	0.054
Intra-epidermal carcinoma	0.9721	0.9732	0.9738	0.9745	4.74	0.192
Actinic Keratosis	0.9668	0.9595	0.9800	0.9807	78.7	< 0.0001
Dysplastic Nevus	0.9606	0.9634	0.9668	0.9663	24.5	< 0.0001

Note: The X^2 test refers to the null hypothesis that there is no difference between the four estimates of the AUROC for each lesion type.

Discussion

The results of this study which is an addition to our other evaluations (10) (9) indicate that the DERM algorithm is capable of detecting skin cancers and potential precursors from images captured by cameras that are in common use, require inexpensive modification and little operator training. The level of accuracy is similar to that of a specialist dermatologist with AUROC ranging from 0.952 for melanoma to .987 for BCC. In addition, given the sequential nature of the algorithm assessment going from most serious (Melanoma) to least serious (Dysplastic Nevus), only 3.2% of the cancers and precursors would not be referred for biopsy or clinical follow-up.

While continued development of the algorithm improved the AUROC for all lesion types, the improvement was statistically significant for three (Melanoma, Actinic Keratosis and Dysplastic Nevus). SCC and BCC approached statistical significance but as both started with an AUROC greater than 0.98, only marginal improvement was possible which we attribute to a ceiling effect given that the upper limit of the AUROC is one.

DERM has improved over four development versions and therefore, has the potential to continue to improve with the addition of more clinical data and refinement of the algorithm. This is one of the strengths of artificial intelligence.

An issue for this study is that we are following the convention of assuming that histopathology is the gold standard against which DERM should be judged. As Claassen pointed out in 2005 this is not intended to imply that the gold standard is without error.(23) For melanoma biopsies two studies show that concordance for melanoma between pathologists is about 75%.(24, 25) It is therefore possible that some of the errors concerning the DERM assessment are because of errors in the gold standard which this study cannot determine.

A limitation of artificial intelligence applications is that their adoption in clinical practice requires much more than a well-developed, validated algorithm. In a recent JAMA 'Viewpoint' Lindsell et al suggested that "A contributing factor [to the slow adoption of AI] is perhaps that model developers and data scientists pay little attention to how a well-performing model will be integrated into health care delivery."(26) A recent small study has shown that dermatology patients are receptive to AI based diagnostic methods for detection of skin cancer within the context of a human physician-patient relationship.(27) This suggests that the reluctance to adopt AI based diagnostic aids is not related to negative patient attitudes. Lindsell et al go on to state that "Designing a useful AI tool in health care should begin with asking what *system change* the AI tool is expected to precipitate."

This study is not able to address all of the issues raised by Lindsell et al but it does allow us to suggest some scenarios where a DERM image analysis can contribute to secondary prevention of skin cancers:

1. Access to specialist dermatology diagnosis is not universal. People who live in rural and remote areas in most countries have access to primary care physicians but as the US Preventive Services Task Force report makes clear, the accuracy of skin cancer detection in primary care is poor and this is supported by the recent Cochrane Review. Access to DERM might mean that fewer patients from remote areas will be identified for distant specialist review and that they would be more likely to have skin cancer that requires specialist care.
2. Prior assessment of lesions by DERM might allow for more accurate triage of patients referred from primary care to secondary dermatology clinics.
3. A third scenario is that DERM would allow a rapid response second opinion for dermatologists in secondary dermatology clinics.

Conclusions

Our study suggests that the use of a trained AI algorithm can be integrated into both primary and secondary care settings in a way that will improve the accuracy of diagnostic skin cancer assessment and reduce the number of unnecessary biopsies referred for histopathology review.

Abbreviations

BCC – Basal cell carcinoma;

SCC – Squamous cell carcinoma;

DERM – Deep Ensemble for the Recognition of Malignancy.

Declarations

Ethics approval and consent to participate

This study did not require patient participation and so no ethics approval was necessary. All observations were made using freely available digital images of skin lesions or healthy skin which had been obtained from consenting adults. Those images which were derived from clinical trials had had ethics approval.

Consent for publication

Not relevant.

Availability of data and materials

The images used for this study were derived from published databases.

Competing interests

JG is an employee of Skin Analytics Ltd.

Funding

The statistical analysis was funded by the Royal Perth Hospital Research Foundation, Perth, Western Australia (https://www.rphresearchfoundation.org.au/). The Foundation had no role in the conduct of the study or the reporting of the results.

Authors' contributions

JG designed the machine learning algorithm and participated in the drafting of the manuscript. MP conducted the statistical analysis and wrote the initial draft of the manuscript.

Acknowledgements

None.

References

1. American Institute for Cancer Research. Melanoma of the skin is the 19th most common cancer worldwide. 2019.
2. U.S. Department of Health and Human Services. The Surgeon General's call to action to prevent skin cancer. Washington, DC: U.S. Dept of Health and Human Services: Office of the Surgeon General; 2014.
3. Khazaei Z., Ghorat F., Jarrahi A. M., Adineh H. A., Sohrabivafa M. GE. Global incidence and mortality of skin cancer by histological subtype and its relationship with the human development index (HDI); an ecology study in 2018. World Cancer Research Journal. 2019;6(e1265).
4. Surdu S, Fitzgerald EF, Bloom MS, Boscoe FP, Carpenter DO, Haase RF, et al. Occupational Exposure to Ultraviolet Radiation and Risk of Non-Melanoma Skin Cancer in a Multinational European Study. PLoS ONE. 2013;8(4).
5. U.S. Department of Health and Human Services. Skin Cancer Prevention Progress Report: 2019. Washington, DC: U.S. Dept of Health and Human Services: Office of the Surgeon General; 2019.
6. Buller DB, Cokkinides V, Hall HI, Hartman AM, Saraiya M, Miller E, et al. Prevalence of sunburn, sun protection, and indoor tanning behaviors among Americans: Review from national surveys and case studies of 3 states. Journal of the American Academy of Dermatology. 2011;65(5, Supplement 1):S114.e1-S.e11.
7. US Preventive Services Task Force. Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2016;316(4):429-35.
8. Weyers W. Screening for malignant melanoma-a critical assessment in historical perspective. Dermatol Pract Concept. 2018;8(2):89-103.
9. Phillips M, Greenhalgh J, Marsden H, Palamaras I. Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy. Dermatol Pract Concept. 2019;10(1):e2020011-e.

10. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, et al. Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions. JAMA Network Open. 2019;2(10):e1913436-e.
11. Skin Analytics. 2020 [Available from: https://skin-analytics.com/. Accessed 8 September 2018.
12. Mendonça T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J, editors. PH2 - A dermoscopic image database for research and benchmarking. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2013 3-7 July 2013.
13. Molemap. 2020 [Available from: https://www.molemap.net.au. Accessed 10 October 2017.
14. Collaboration TISI. 2020 [Available from: https://www.isic-archive.com/. Accessed 10 October 2017.
15. Giuseppe Argenziano, H. Peter Soyer, Vincenzo De Giorgio, Domenico Piccolo, Paolo Carli, Mario Delfino, Angela Ferrari, Rainer Hofmann-Wellenhof, Daniela Massi, Giampiero Mazzocchetti, Massimiliano Scalvenzi, Ingrid H. Wolf, MDInteractive Atlas of Dermoscopy, Milan, Italy, 2000, Edra Medical Publishing and New Media. ISBN 88-86457-30-8.
16. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data. 2018;5(1):180161.
17. Dermnet. 2020 [Available from: http://www.dermnet.com/. Accessed 10 October 2017.
18. Collins GS, Reitsma JB, Altman DG, Moons GM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Annals of Internal Medicine. 2015;162:55-63.
19. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of residual Connections on learning 2016 [cited 2016. Available from: arXiv.1602 07261.
20. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press;: Oxford University Press Inc, NY; 2003.
21. Kirby JS, Scharnitz T, Seiverling EV, Ahrns H, Ferguson S. Actinic Keratosis Clinical Practice Guidelines: An Appraisal of Quality. Dermatology Research and Practice. 2015;2015:456071.
22. Nahhas AF, Scarbrough CA, Trotter S. A Review of the Global Guidelines on Surgical Margins for Nonmelanoma Skin Cancers. J Clin Aesthet Dermatol. 2017;10(4):37-46.
23. Claassen JAHR. The gold standard: not a golden standard. BMJ. 2005;330(7500):1121.
24. Corona R, Mele A, Amini M, De Rosa G, Coppola G, Piccardi P, et al. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. Journal of Clinical Oncology. 1996;14(4):1218-23.
25. Lodha S, Saggar S, Celebi JT, Silvers DN. Discordance in the histopathologic diagnosis of difficult melanocytic neoplasms in the clinical setting. Journal of Cutaneous Pathology. 2008;35(4):349-52.
26. Lindsell CJ, Stead WW, Johnson KB. Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem. JAMA. 2020;323(21):2141-2.
27. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. JAMA Dermatology. 2020;156(5):501-12.

Figures

Figure 1

Sequence for cycling lesion assessment