Artificial Intelligence and Skin Cancer Detection

Laura K. Ferris, MD, PhD
Professor of Dermatology
University of Pittsburgh

UPMC LIFE CHANGING MEDICINE

Artificial (augmented) intelligence and skin cancer diagnosis

- Applied to images of skin lesions
  - Can use on clinical or dermatoscopic images (or others like RCM)
  - Can also be used on non-image-based data (genetic profile)

- Basic principles:
  - Need training set of images (known diagnosis/benign vs malignant)
  - Computer is trained to differentiate lesions in training set
    - CNN– computer picks the features based on multiple layers of processing – can uncover features not identified by humans
  - Validate on a non-overlapping test set of images
  - Computer will give a score or probability of malignancy

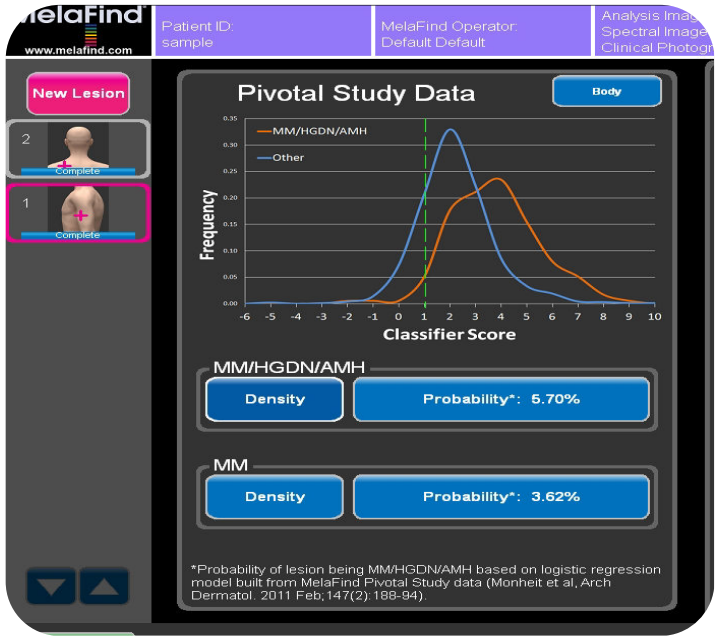
J Am Acad Dermatol. 2022 Dec;87(6):1336-1342. J Clin Med. 2022 Nov 18;11(22):6826.

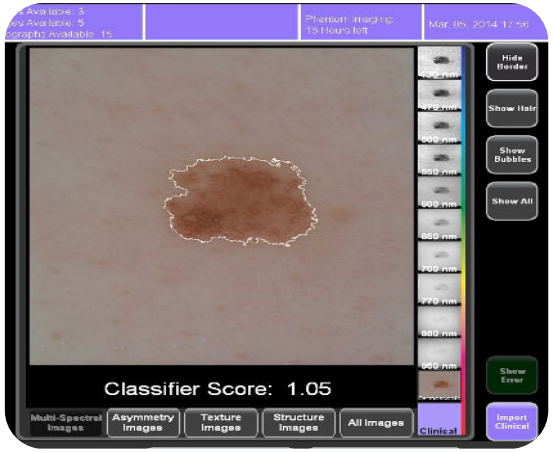
Human vs Machine – key differences in determining if a lesion is benign or malignant

Parameter	Human	Machine
Context	Other lesions, patient history, risk factors	Usually evaluates lesion in isolation
Scope	Can evaluate all skin surfaces	Can only evaluate the lesion a person chooses to use it on
Objectivity	• Shaded by experience, patient's degree of concern, own fear of missing melanoma, incentive to biopsy or not • Tend to favor biopsy of benign over missing malignant	Objective: Can set a threshold score to maximize sensitivity at cost of specificity
Learning	Learn incrementally one patient at a time, one paper at a time, years of training	Can train classifier or scoring system on a set of images in hours / days
Features evaluated	Can explain "why" (ABCDE, dermatoscopic features)	Cannot always explain "why": Identify and use new features ; process large amounts and layers of data

MELAFIND- What can we learn?

- Sensitivity comes at the cost of specificity
- Limited utility: Recommends biopsy of about 90% of lesions
- Expensive optics/ machine with high upfront cost limit use
- Fixed classifier- cannot "learn" in real time →need FDA reapproval
- Product no longer available / supported





	Sensitivity	Specificity
MelaFind	97%	9%
Readers	72%	51%

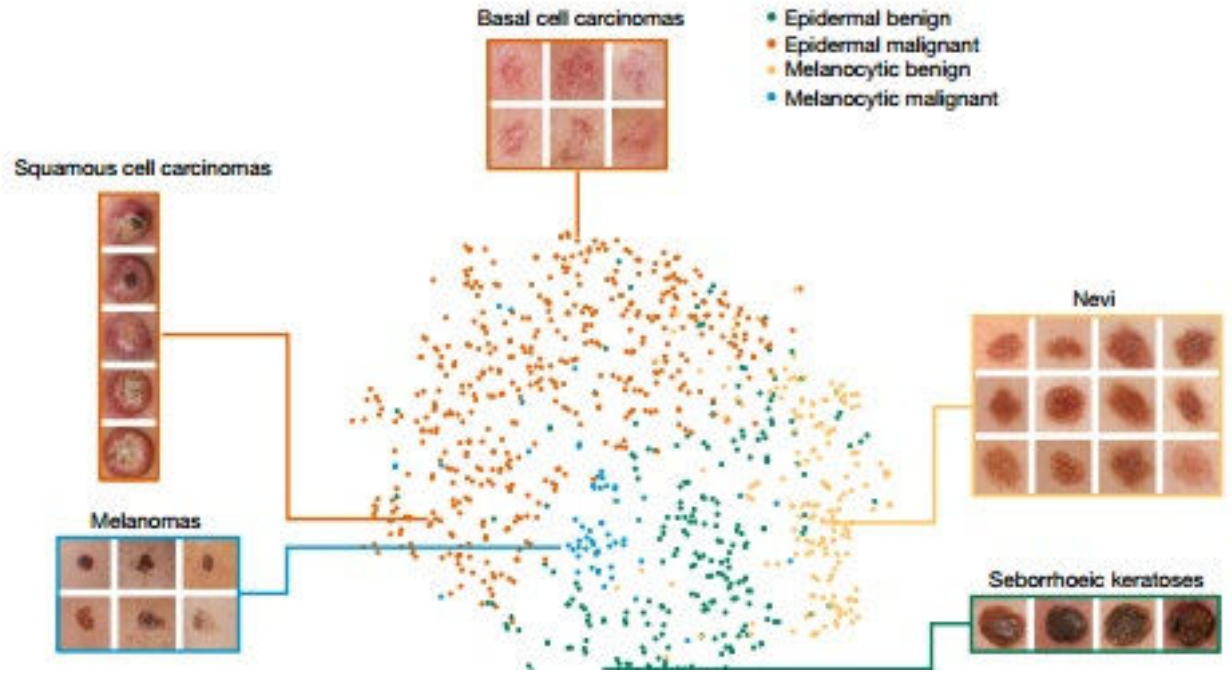
Threshold	Sensitivity (95% CI)	Specificity (95% CI)
≥ -3	100.0% (97.9-100%)	0.8% (0.4-1.4%)
≥ -2	99.4% (96.9-100%)	1.3% (0.8-2.0%)
≥ -1	98.9% (95.9-99.9%)	3.6% (2.7-4.7%)
≥ 0	98.3% (95.1-99.6%)	10.8% (9.2-12.5%)
≥ 1	93.1% (88.3-96.4%)	29.8% (27.4-32.2%)
≥ 2	75.4% (78.4-81.6%)	60.0% (57.4-62.5%)
≥ 3	54.3% (46.6-61.8%)	81.8% (79.7-83.8%)
≥ 4	30.9% (24.1-38.3%)	91.4% (89.9-92.8%)
≥ 5	15.4% (10.4-21.7%)	96.1% (95.0-97.0%)
≥ 6	7.4% (4.0-12.4%)	98.8% (98.2-99.4%)

Monheit et al, Arch Dermatol. 2011 Feb;147(2):188-94. MelaFind Package insert

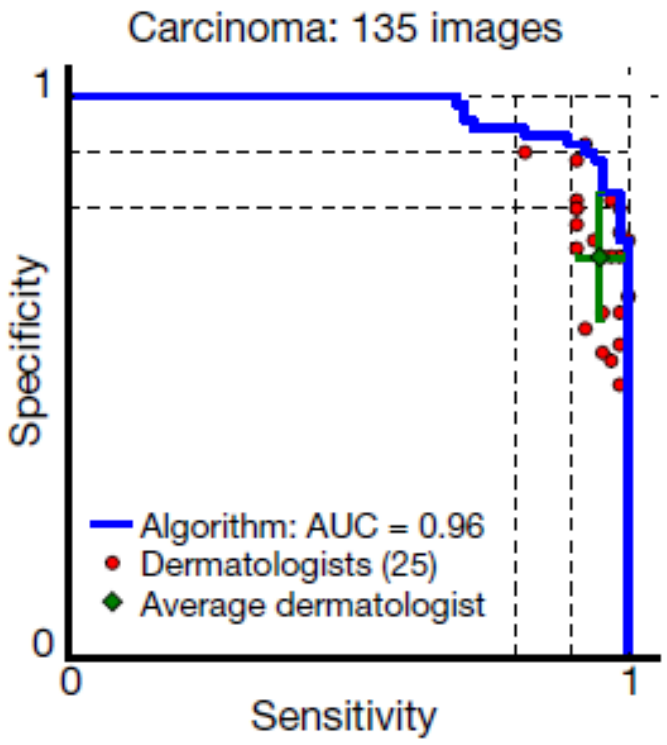
Dermatologist-level classification of skin cancer with deep neural networks

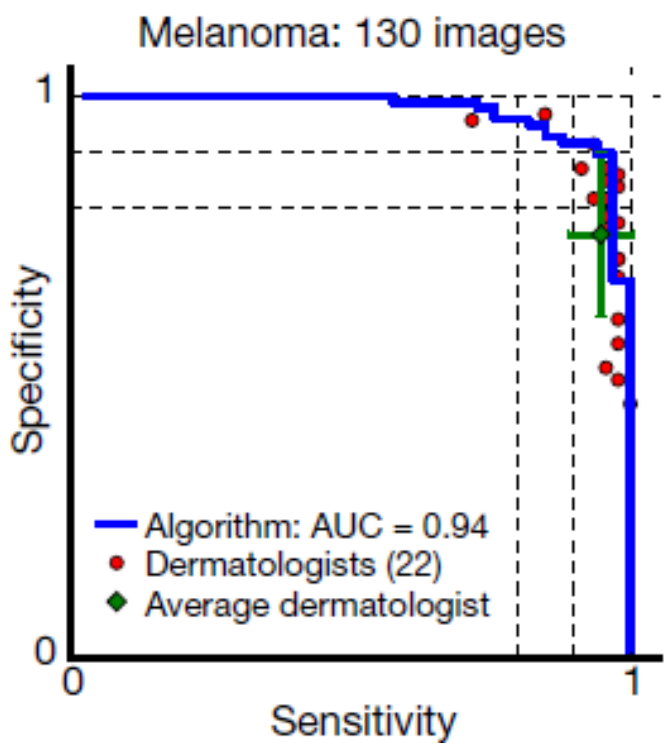
- Trained on >100,000 skin lesion images of >2,000 diseases
- Compared accuracy to dermatologists
- Computer
  - Melanocytic test: Melanoma vs nevi
  - Epidermal test: BCC / SCC vs SK
  - NOT SK vs melanoma!
- Dermatologists : would you biopsy this (Y/N)?

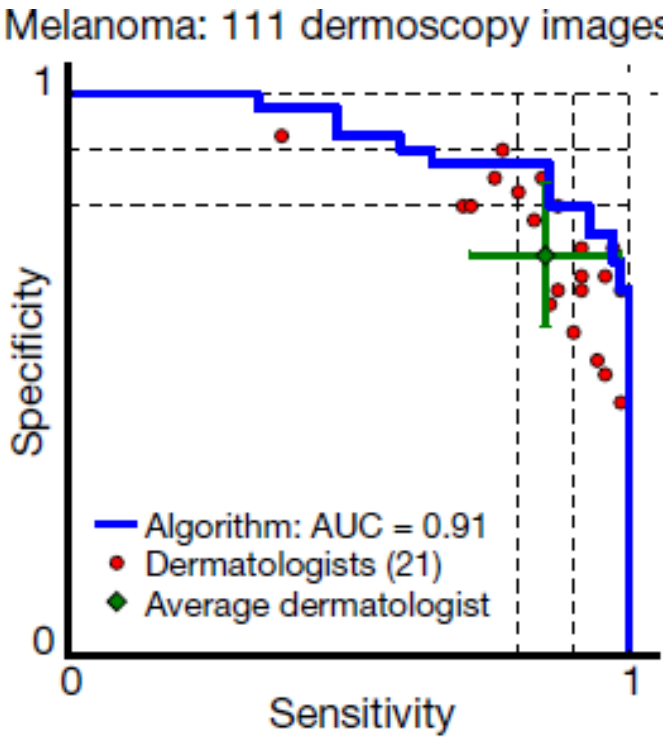
Clinical or dermatoscopic images



Nature. 2017 Feb 2;542(7639):115-118.







33 KCs, 70 SKs

33 melanomas, 97 benign nevi

71 melanomas, 40 benign nevi

Red dot above the curve = Dermatologist outperformed computer

Red dot below the curve = Computer outperformed the dermatologist

No data on lesion thickness

Worse performance with dermoscopy → harder images?

Nature. 2017 Feb 2;542(7639):115-118.

Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study

Philipp Tschandl, Noel Codella, Bengü Nisa Akay, Giuseppe Argenziano, Ralph P Braun, Horacio Cabo, David Gutman, Allan Halpern, Brian Helba, Rainer Hofmann-Wellenhof, Aimilios Lallas, Jan Lapins, Caterina Longo, Josep Malvehy, Michael A Marchetti, Ashfaq Marghoob, Scott Menzies, Amanda Oakley, John Paoli, Susana Puig, Christoph Rinner, Cliff Rosendahl, Alon Scope, Christoph Sinz, H Peter Soyer, Luc Thomas, Iris Zalaudek, Harald Kittler

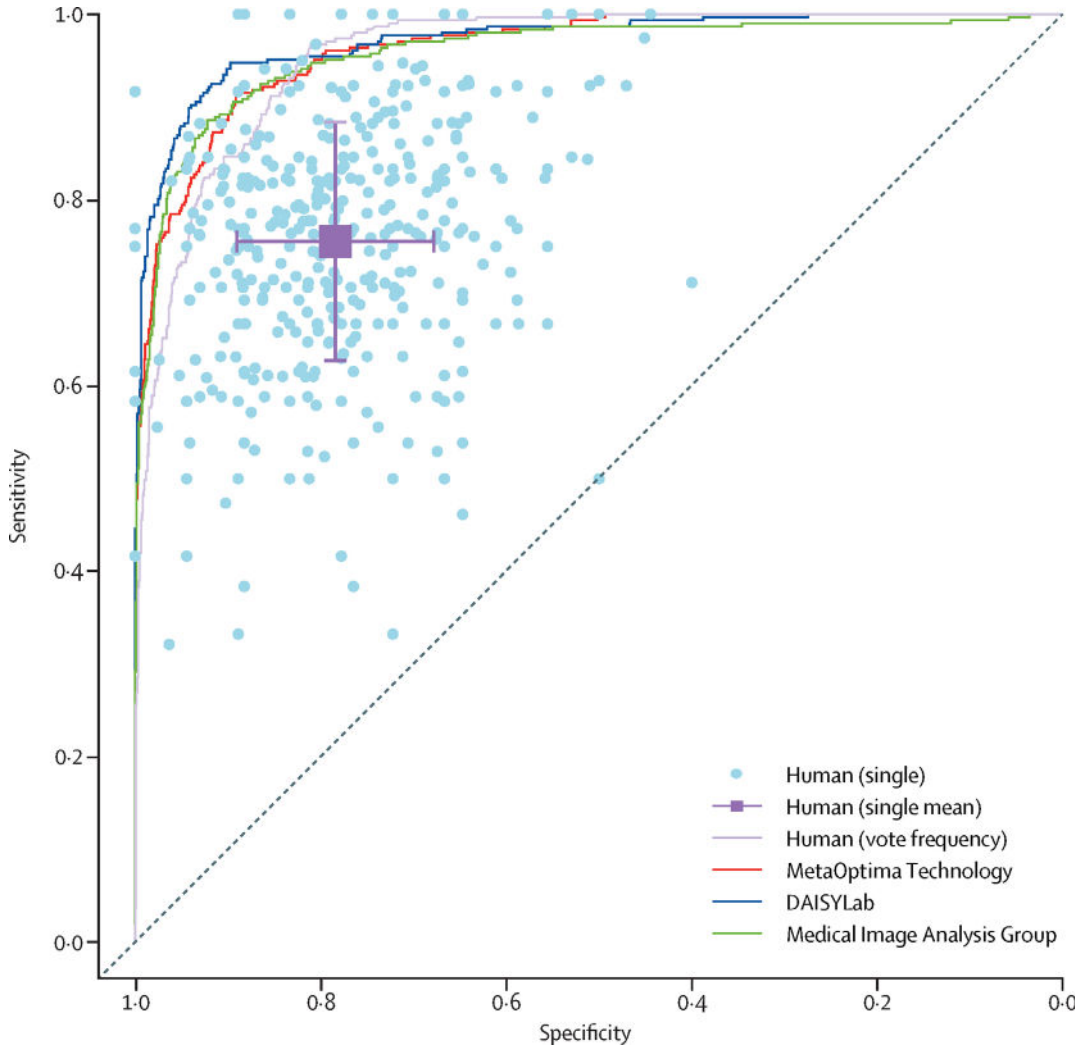
- **Diagnoses from human readers were compared with those of 139 algorithms created by 77 machine-learning labs, who participated in the International Skin Imaging Collaboration 2018 challenge**

Lancet Oncol. 2019 Jul;20(7):938-947.

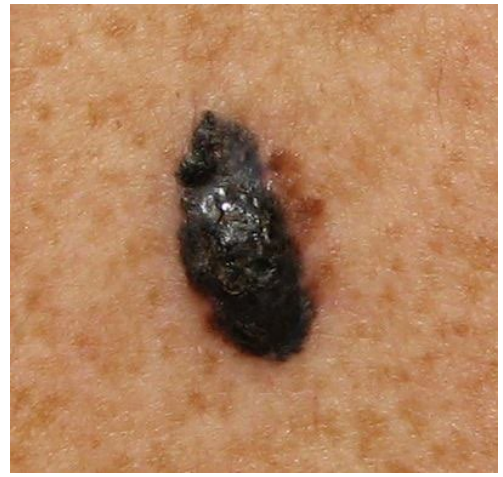
	Sensitivity	Specificity	Negative predictive value	Positive predictive value
All human readers (n=511)				
Melanoma	73·1% (65·8–79·6)	92·8% (91·3–94·2)	96·4% (95·3–97·4)	56·6% (49·7–63·2)
Expert readers (n=27)				
Melanoma	67·8% (60·3–74·8)	94·0% (92·5–95·2)	95·8% (94·6–96·8)	58·9% (51·7–65·8)
All algorithms (n=139)				
Melanoma	67·3% (59·7–74·2)	97·0% (96·0–97·9)	95·9% (94·7–96·9)	74·2% (66·6–80·9)
Top three algorithms (n=3)				
Melanoma	81·9% (75·4–87·3)	96·2% (95·1–97·2)	97·6% (96·7–98·4)	73·6% (66·9–79·6)

Top 3 algorithms also performed equivalent to humans on SKs

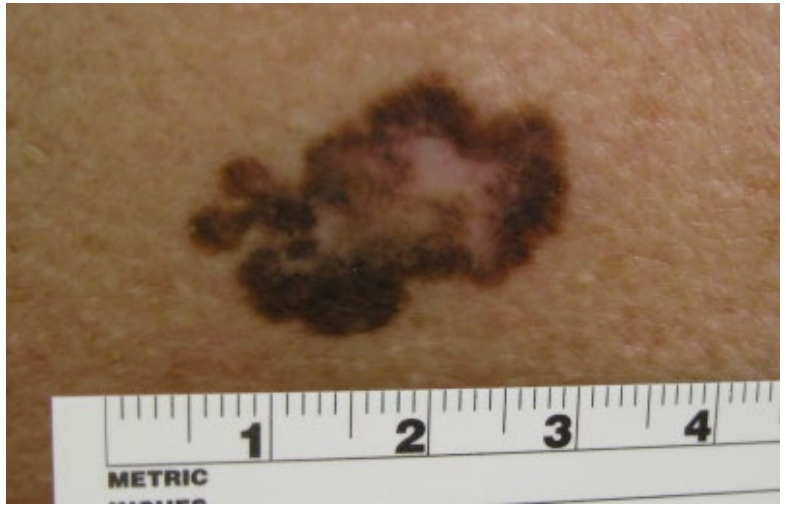
Discrimination of malignant from benign pigmented skin lesion

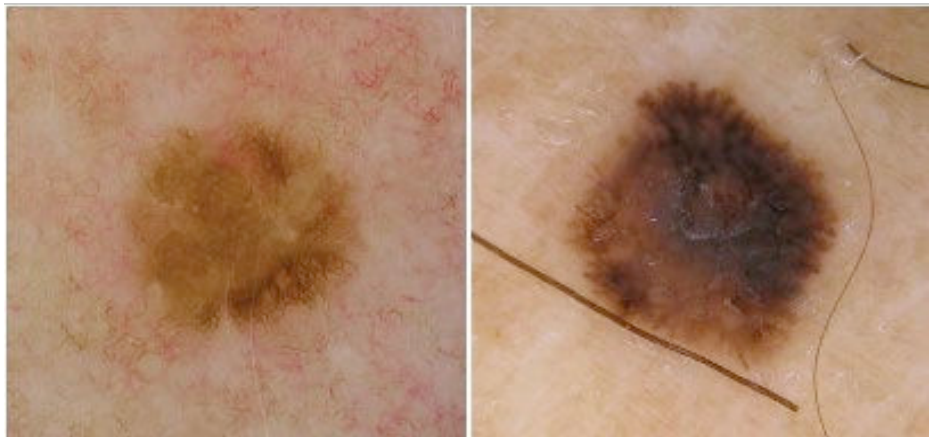


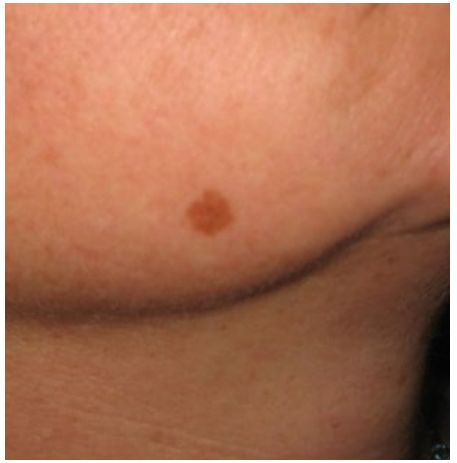
Comparing human vs. machine with reader studies



- Factors that will bias results
  - Ratio of melanomas to total lesions
  - Amount of information given to readers (clinical history, clinical image, dermatoscopic image)
  - Image quality
  - Difficulty of lesions– Reader studies helpful for this!
  - Stakes are different on a computer vs. patient

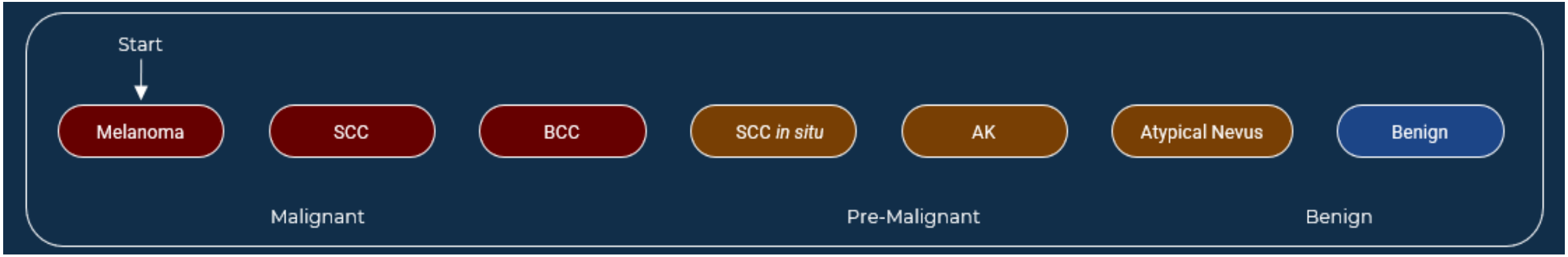


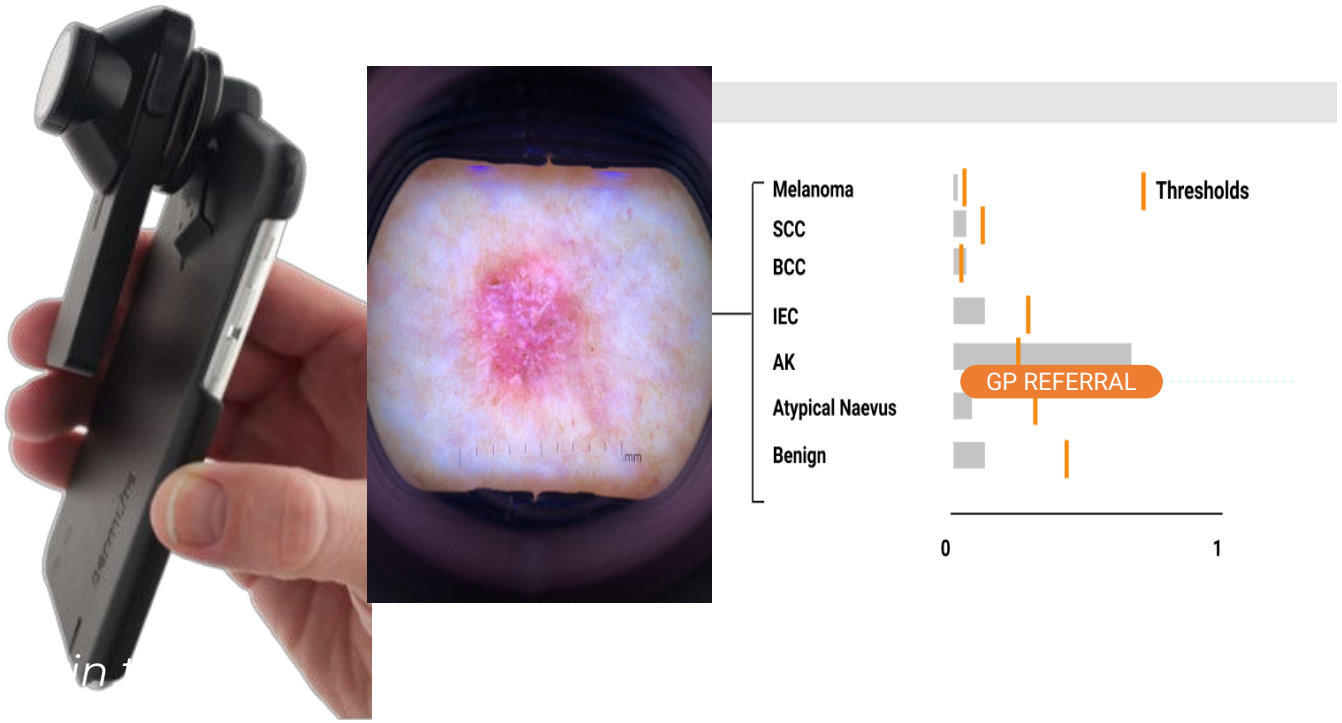




Reader Studies			
Marchetti (2017)	Dermoscopy only	100 Lesions, 50 Melanomas	82% sensitivity/ 59% specificity
Ferris (2017)	Clinical image and history, dermoscopy	60 lesions, 8 melanomas	95% Sensitivity/ 32.1% specificity
Ferris (2015)	Dermoscopy only	65 lesions, 25 melanomas	70.8% Sensitivity/ 58.7% specificity
Monheit (2011)	Clinical images and history, dermoscopy	50 lesions, 25 melanomas	78% Sensitivity/ Specificity NR
Nevisense (FDA report 2017)	Clinical images and dermoscopy	Lesion/melanoma # NR	77.2% Sensitivity/ 53.1% specificity

Deep Ensemble for the Recognition of Malignancy (DERM)





- Images provided by SkinAnalytics

DERM Accuracy (malignant vs benign)

	Target	DERMv3 15 Jul 2021 - 21 Apr 2022	DERMv3.0.1+ 22 Apr 2022 - 19 Aug 2022
Melanoma	95%	97.20% (139/143)	100.00% (47/47)
SCC	95%	99.38% (159/160)	100.00% (44/44)
BCC	90%	99.33% (447/450)	100.00% (69/69)
Malignant 'Other'		70.00% (7/10)	100.00% (1/1)
All skin cancer		98.56% (752/763)	100.00% (161/161)
Bowen's disease (IEC)	90%	95.69% (222/232)	96.94% (95/98)
Actinic Keratosis	90%	93.58% (510/545)	92.59% (175/189)
Atypical Naevus		76.10% (191/251)	60.67% (91/150)
Refer 'Other'		77.78% (14/18)	100.00% (1/1)
All pre-malignant		93.84% (746/795)	94.10% (271/288)
Benign *Specificity*		49.00% (1,442/2,943)	71.68% (987/1,377)

- SkinAnalytics, unpublished data

Elastic Scattering Spectroscopy (ESS)

- Handheld device that measures spectra of skin lesions and uses CNN to classify the lesion's properties against those of known malignant and benign lesions
- Trial in 1005 patients/ 1579 lesions, all evaluated by PCPs and biopsied (Merry et al)
  - 48 melanomas, 90 BCC, 86 SCC, 1355 benign

PCP sensitivity = 83.0%
ESS device sensitivity = 95.5% (p<0.0001)
ESS device specificity = 20.7%

NPV of device = 96.6%
PPV =16.6%

Concordance between PCP and biopsy		
PCP diagnosis	Biopsy diagnosis	
	Benign	Malignant
Benign	734 (54.2%)	38 (17.0%)
Malignant	621 (45.8%)	186 (83.0%)
Total	1355	224

Concordance between device and biopsy		
Device	Biopsy diagnosis	
	Benign	Malignant
Benign	281 (20.7%)	10 (4.5%)
Malignant	1074 (79.3%)	214 (95.5%)
Total	1355	224

- In another study (Hartman et al) of lesions biopsied by dermatologists for suspicion of melanoma
  - (44 melanomas, 326 benign nevi),
  - Device sensitivity = 95.5%; specificity= 32.5%; NPV=98%; PPV=47%

- Poster, Merry et. al., Maui Dermatology Meeting 2023; Poster, Hartman et al. AAD 2022 Annual Meeting;

JAMA Dermatology | Consensus Statement

Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology

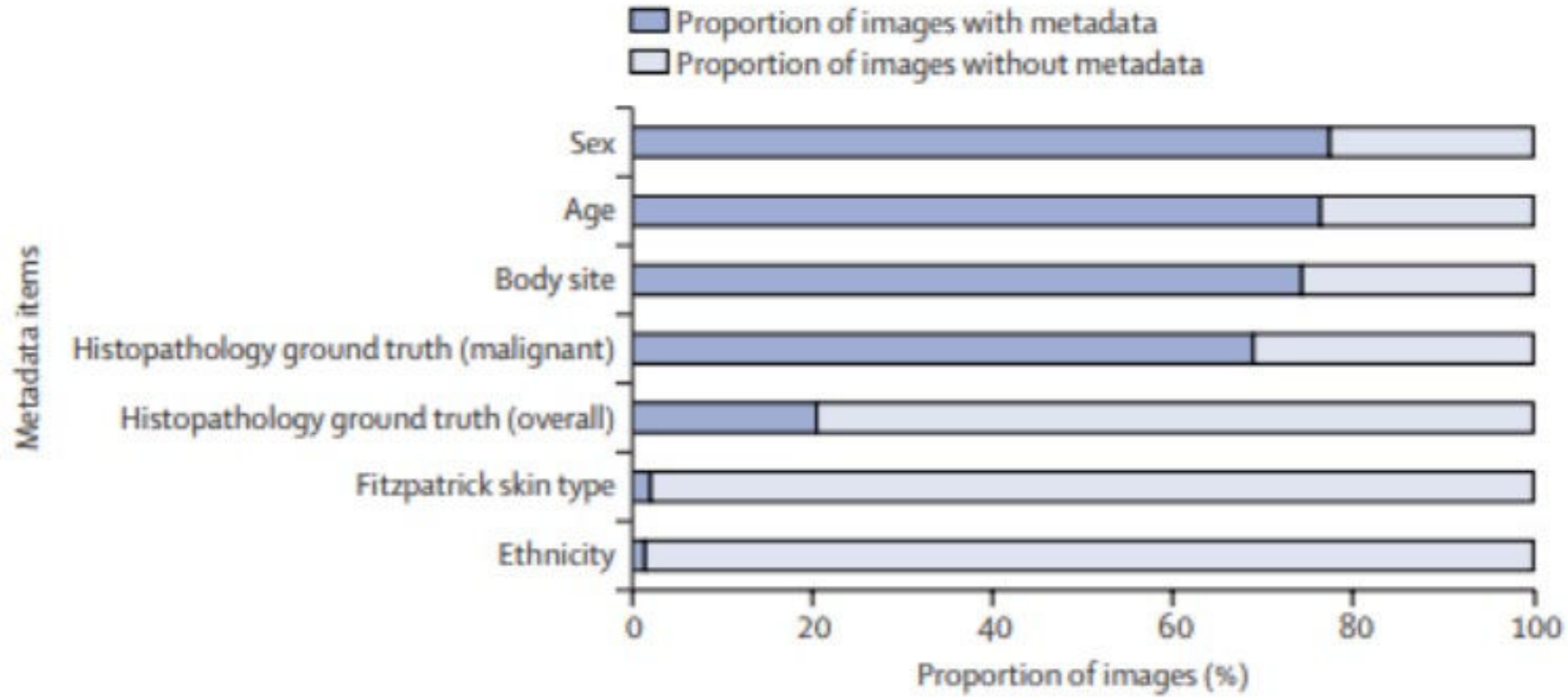
CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group

Guidelines intended to address challenges in dermatology image-based AI that hinder clinical translation, including

- **lack of image standardization**
- concerns about potential sources of **bias**
- factors that cause performance degradation

JAMA Dermatol. 2022 Jan 1;158(1):90-96.

Is AI equally applicable to all patients?



Proportion of open access skin images (n=106950 images in 21 datasets) reporting image and subject metadata.

- Lancet Digit Health. 2022 Jan;4(1):e64-e74. Epub 2021 Nov 9. PMID: 34772649

STUDY

Diagnostic Inaccuracy of Smartphone Applications for Melanoma Detection

JAMA Dermatol. 2013 Apr;149(4):422-6.

4 smart phone apps for melanoma detection

- 60 melanoma, 128 benign lesions
- All confirmed by dermatopathologist
- App 1-3 = automated, cheap, instant
- App 4 = S/F teledermatology

Table 2. Sensitivity and Specificity of Applications Using Evaluable Images

Application No.	Evaluable Image, No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
1	182 (96.8)	70.0 (56.6-80.8)	39.3 (30.7-48.6)
2	185 (98.4)	69.0 (55.3-80.1)	37.0 (28.7-46.1)
3	170 (90.4)	6.8 (2.2-17.3)	93.7 (87.0-97.2)
4	159 (84.6)	98.1 (88.8-99.9)	30.4 (22.1-40.3)

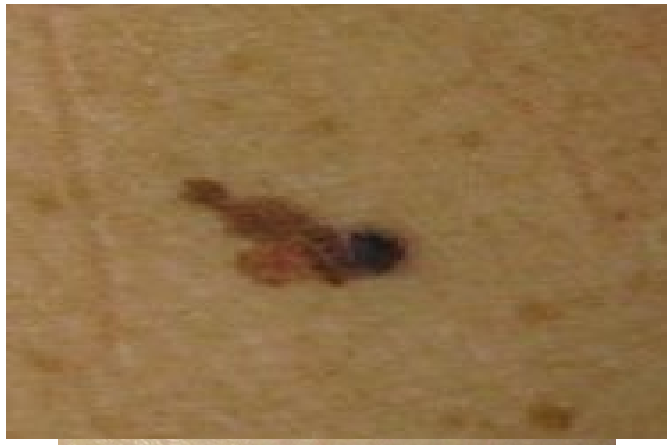
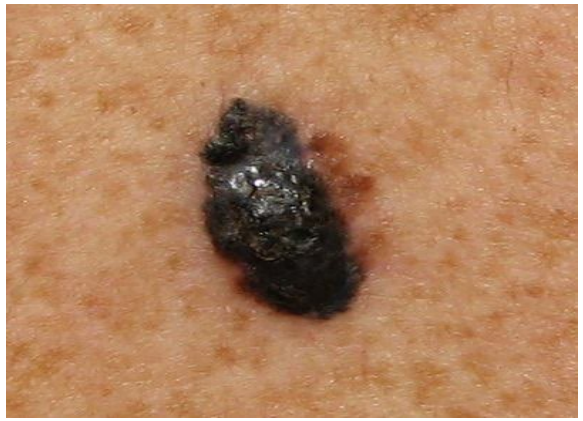
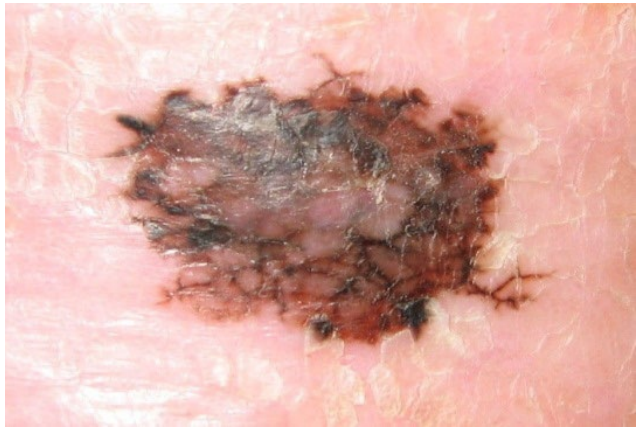
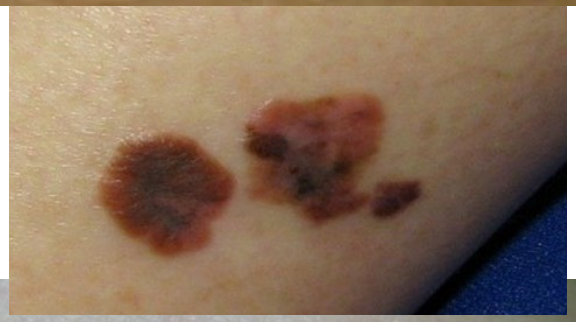
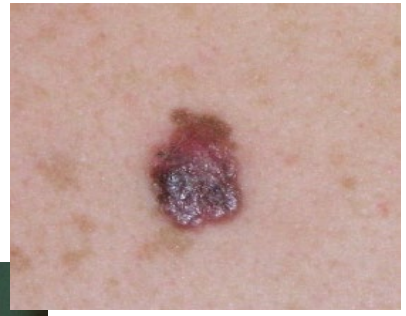
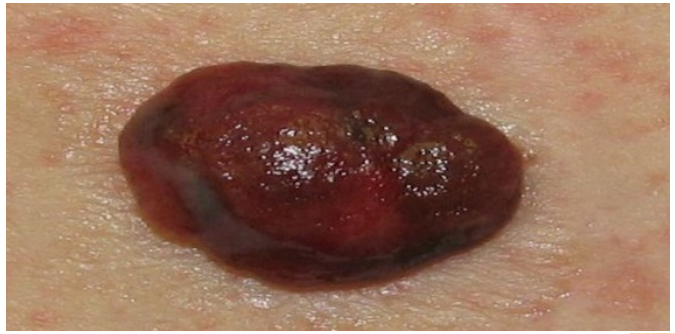
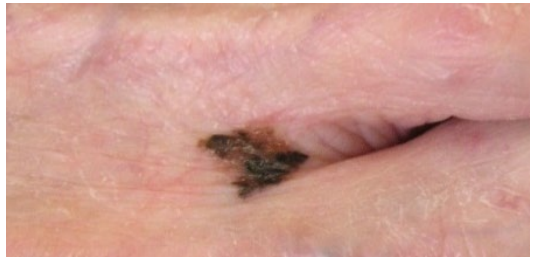
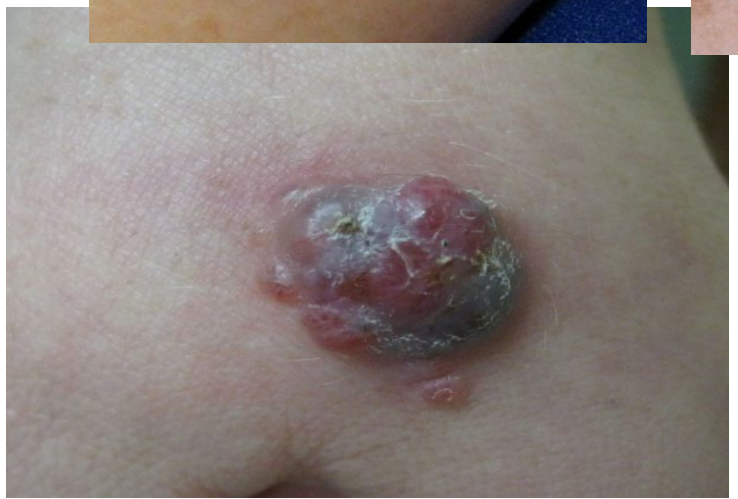
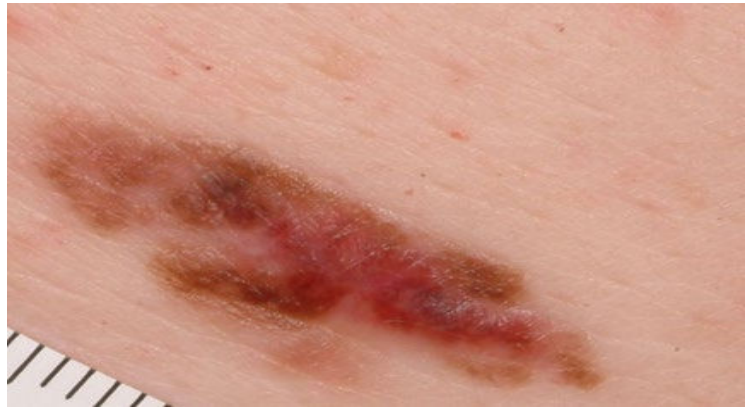
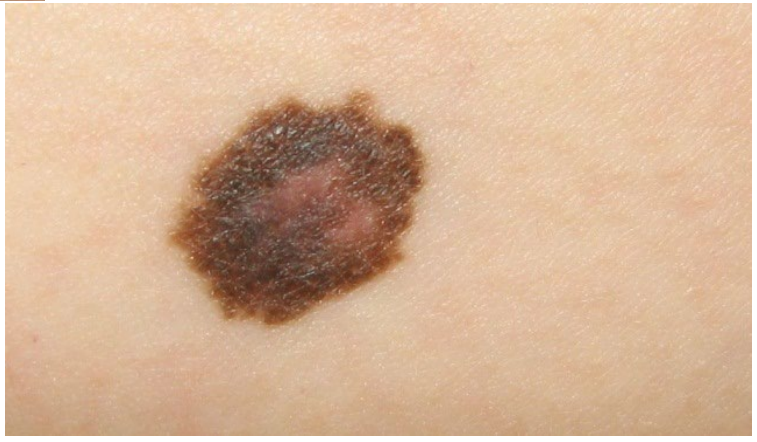
Accuracy of commercially available smartphone applications for the detection of melanoma

Br J Dermatol. 2022 Apr;186(4):744-746.

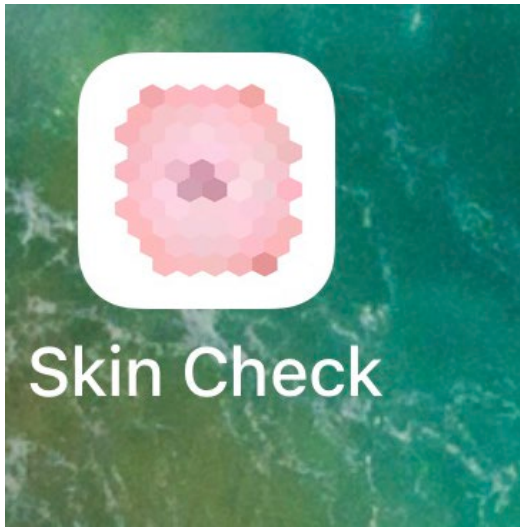
Evaluation of 25 apps :
15 invasive melanomas/ 15 benign with histology
5 benign clinically in SOC patients:

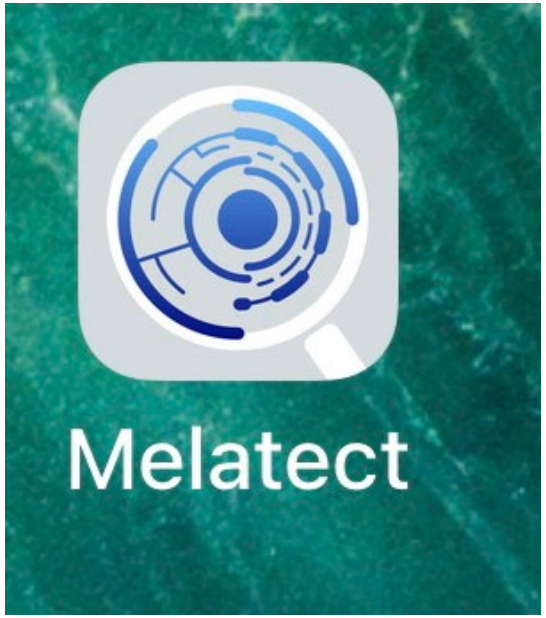
- mean sensitivity = 0.28 (95% CI 0.17–039)
- mean specificity = 0.81 (95% CI 0.71–091)
- mean accuracy = 0.59 (95% CI 0.55–062)

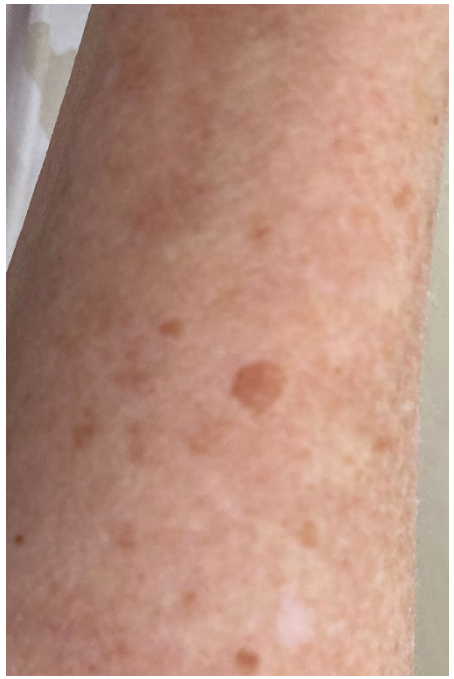
Missed melanomas

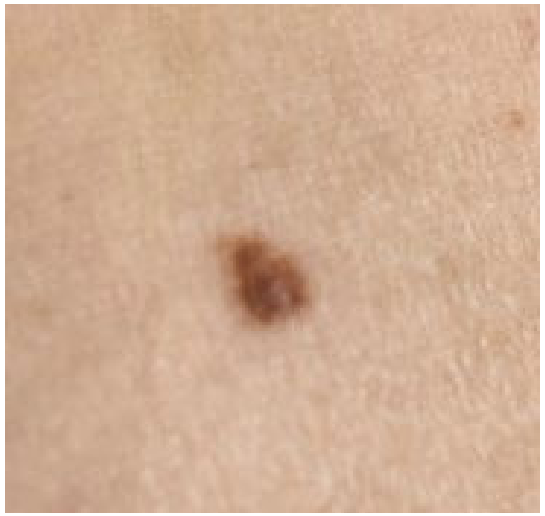











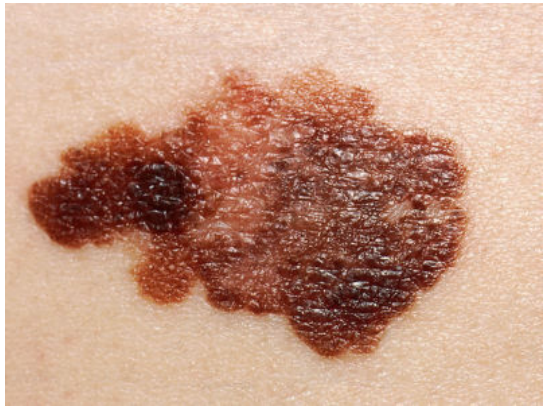

scanoma-i...


Skin Check


Melatect

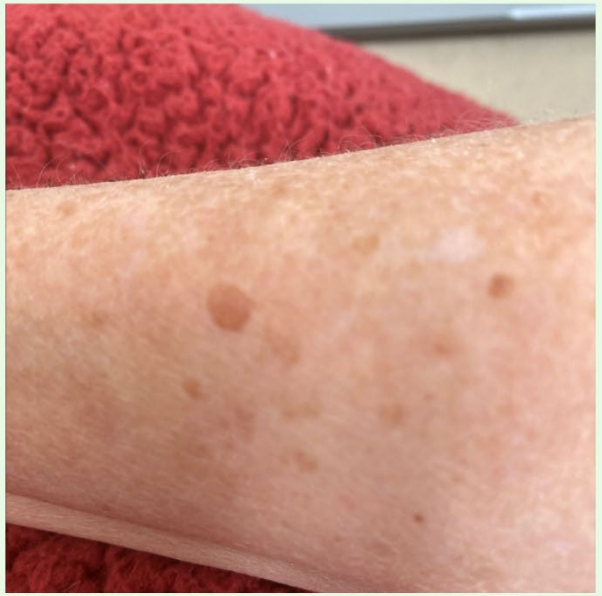


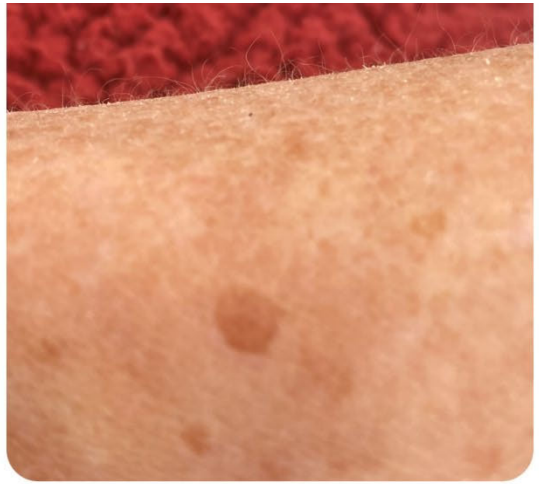




The mole present may be benign

We recommend you contact a dermatologist for an evaluation if concerns about your mole persist. If you believe there has been a mistake, please contact our team.





Not Concerning: 93%

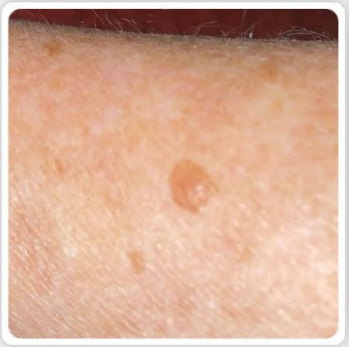
Save Analysis?

Body Tag:

Examples of benign moles used to train classifier:

Save Analysis

AI Results

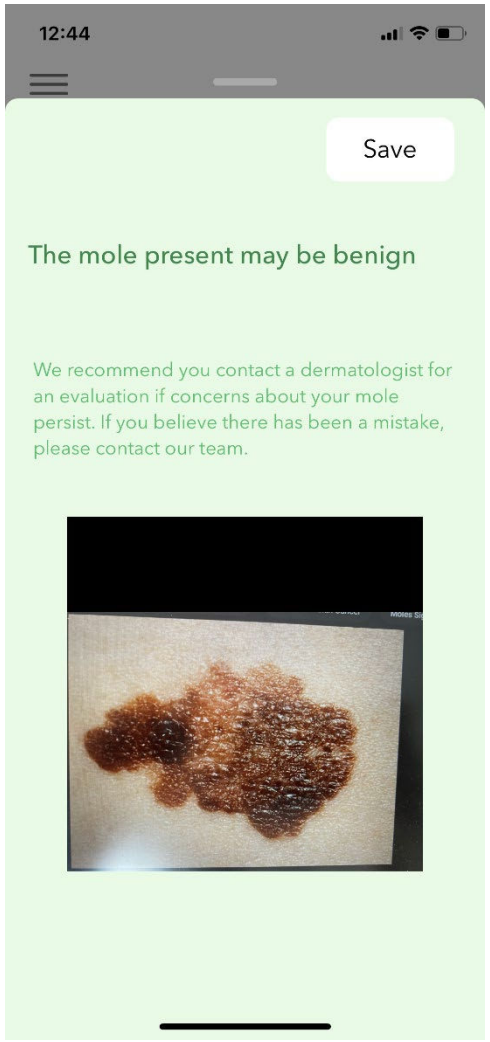


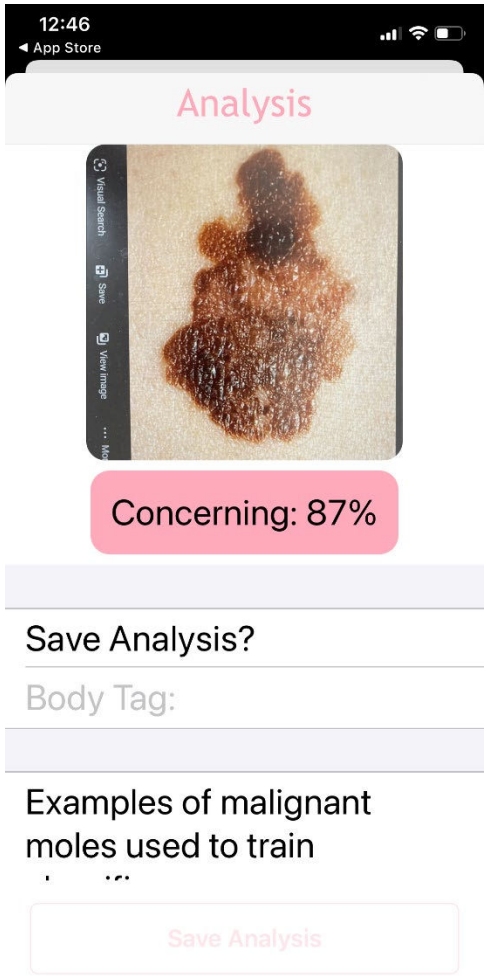
71% NOT CONCERNING

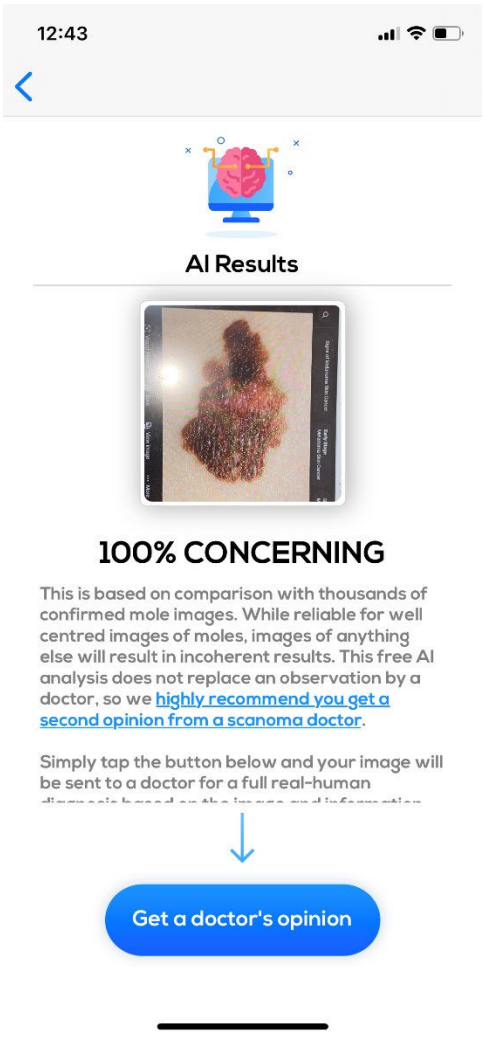
This is based on comparison with thousands of confirmed mole images. While reliable for well centred images of moles, images of anything else will result in incoherent results. This free AI analysis does not replace an observation by a doctor, so we highly recommend you get a second opinion from a scanoma doctor.

Simply tap the button below and your image will be sent to a doctor for a full real-human

Get a doctor's opinion

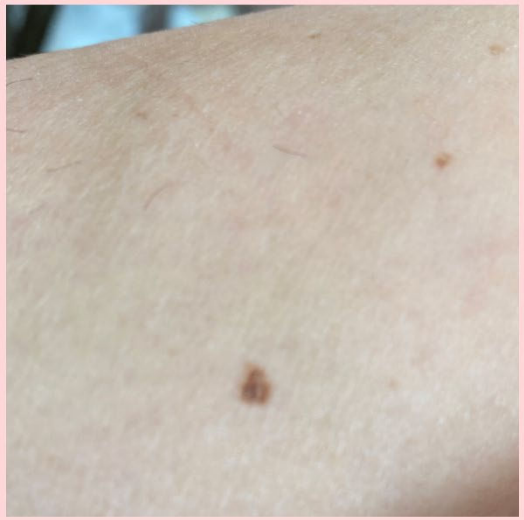

12:44
Save
The mole present may be benign
We recommend you contact a dermatologist for an evaluation if concerns about your mole persist. If you believe there has been a mistake, please contact our team.


12:46
App Store
Analysis
Visual Search
Save
View image
Concerning: 87%
Save Analysis?
Body Tag:
Examples of malignant moles used to train
Save Analysis

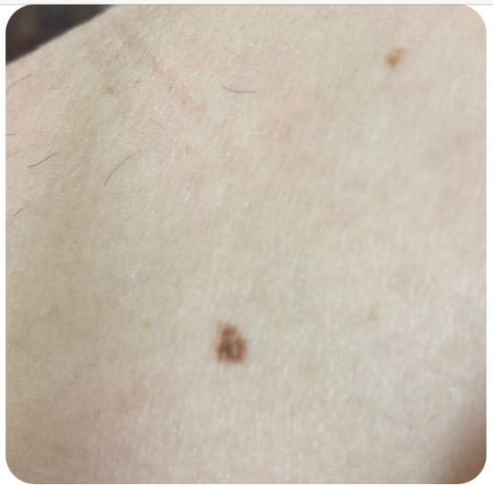

12:43
AI Results
100% CONCERNING
This is based on comparison with thousands of confirmed mole images. While reliable for well centred images of moles, images of anything else will result in incoherent results. This free AI analysis does not replace an observation by a doctor, so we highly recommend you get a second opinion from a scanoma doctor.
Simply tap the button below and your image will be sent to a doctor for a full real-human
Get a doctor's opinion

The mole present may be malignant

We reccomend contacting your health care provider immediately or getting in touch with a specialist using the MyDoctors page tab. If you believe there has been a mistake, please contact our team.



Analysis



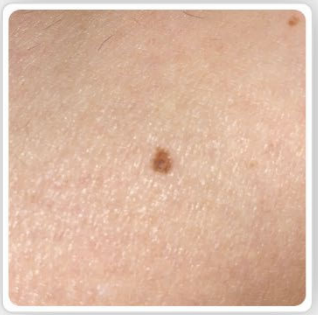
Not Concerning: 96%

Save Analysis?

Body Tag:

Examples of benign moles used to train classifier:

AI Results



69% CONCERNING

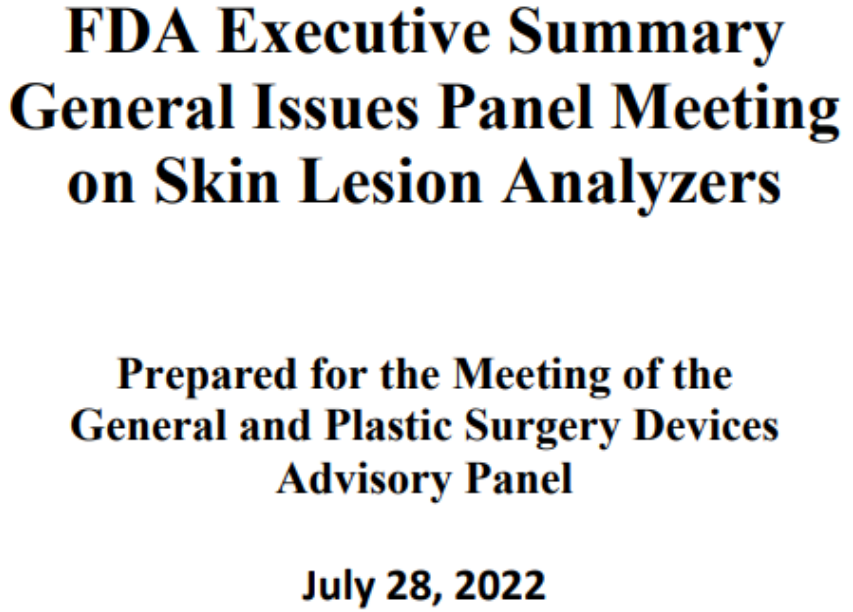
This is based on comparison with thousands of confirmed mole images. While reliable for well centred images of moles, images of anything else will result in incoherent results. This free AI analysis does not replace an observation by a doctor, so we highly recommend you get a second opinion from a scanoma doctor.

Simply tap the button below and your image will be sent to a doctor for a full real-human

Get a doctor's opinion

FDA oversight of AI devices

FDA device classifications	Example
Class I (low risk)	dermatoscope
Class II (moderate risk)	Phototherapy unit
Class III (high risk)	Soft tissue fillers



- Two currently approved skin cancer detection devices for use by dermatologists
  - Melafind
  - Nevisense (EIS device)

Class III devices: require full premarket approval (PMA)

- Proposal to change classification to class II device: Opposed by makers of Nevisense and by AAD
- Discussed what is needed as “ground truth” for validation studies: pathology for all vs. dermatologist consensus diagnosis for benign lesions

- www.fda.gov

Examples of Device Software Functions the FDA Regulates

Software functions (typically mobile apps) that transform a mobile platform into a regulated medical device and therefore are the focus of the FDA's regulatory oversight:

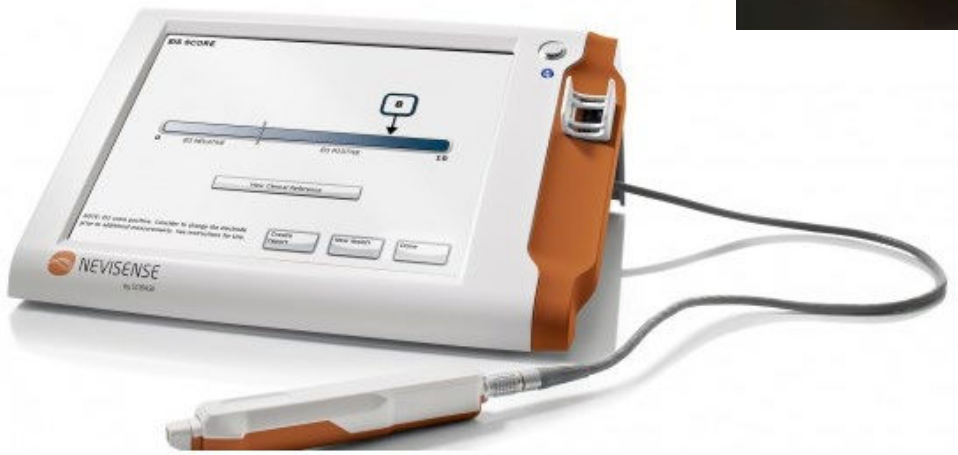
> “Analyze an image of a skin lesion using mathematical algorithms, such as fractal analysis, and provide the user with an assessment of the risk of the lesion.”

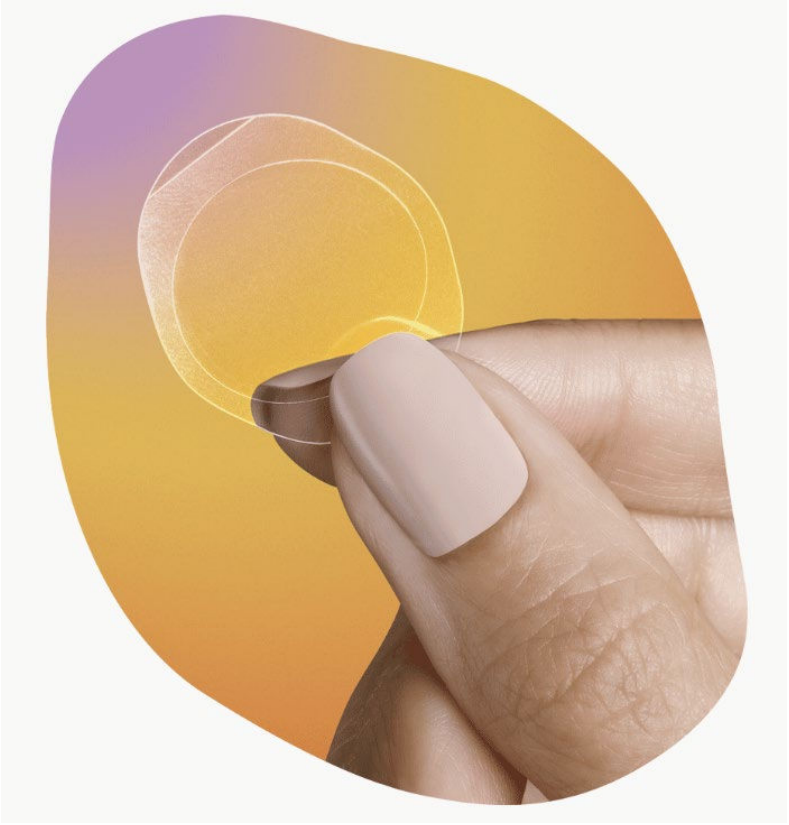
Examples of Device Software Functions the FDA Regulates | FDA (updated 09/29/2022)

Regulation of melanoma diagnostic tools: same output, different levels of scrutiny

- Devices are regulated by the FDA → Gene expression test are not!

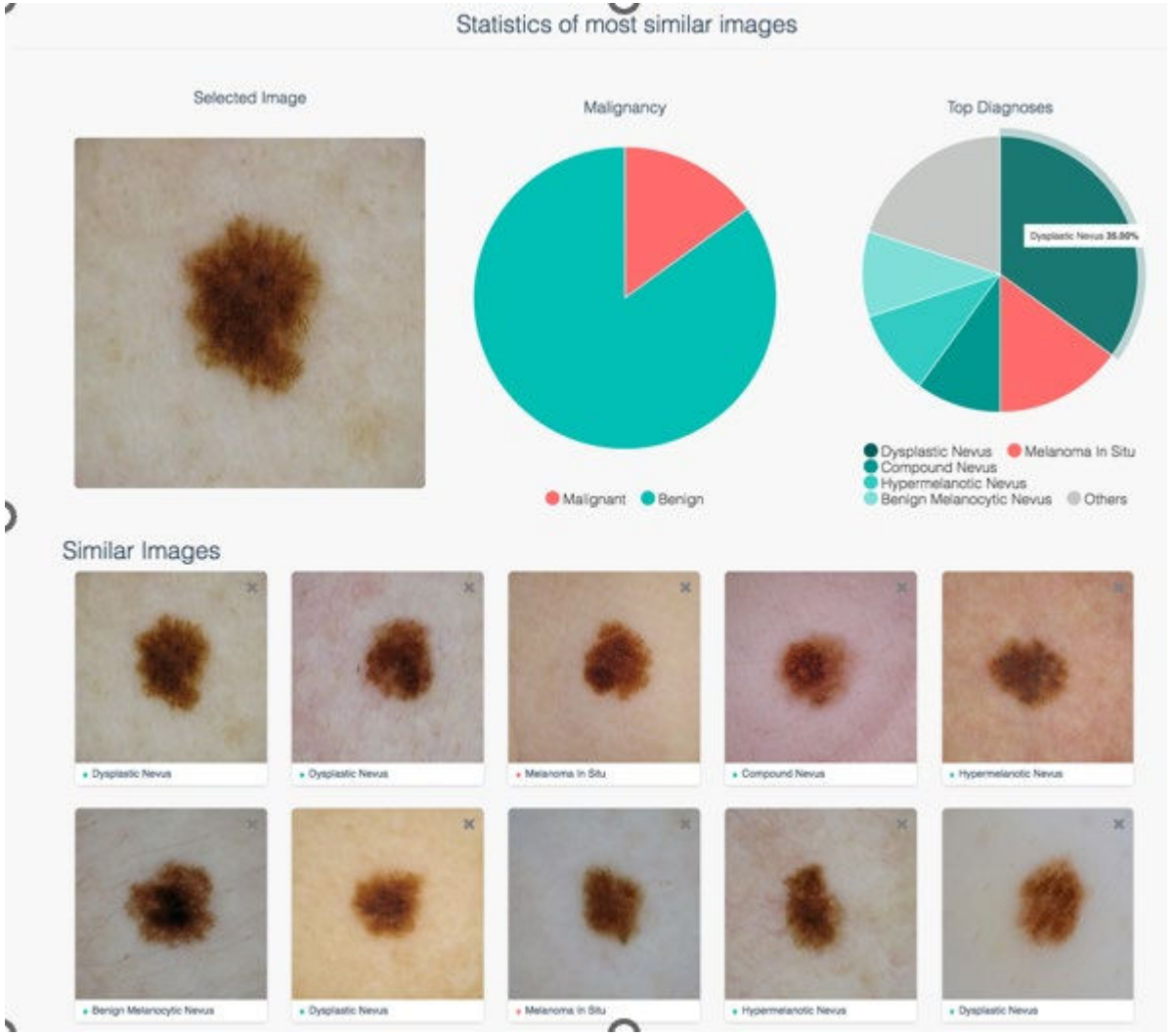
Gene expression profiling tests just need to have CLIA certification for the labs in which they are run





- www.dermtech.com (accessed Jan 11, 2023); MELANOMA – Scibase (accessed Jan 11, 2023) Research - Skin Analytics (skin-analytics.com) (accessed Jan 11, 2023)

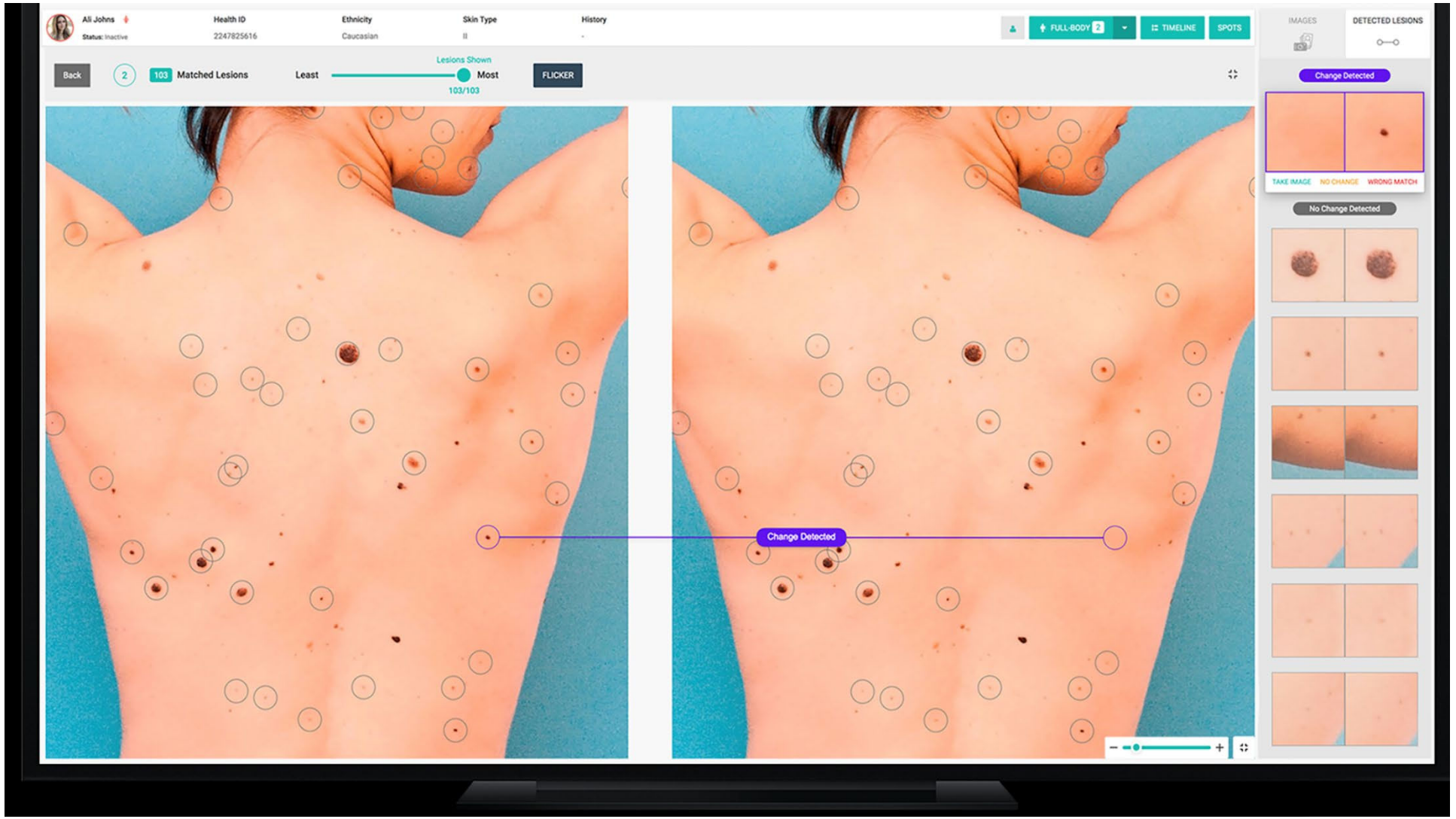
Using AI to rapidly search a database to find similar images



- User decides which images look most like one being evaluated and if a biopsy is warranted
- Like an atlas / textbook
- Not giving a diagnosis or making a treatment recommendation → **not a medical device**

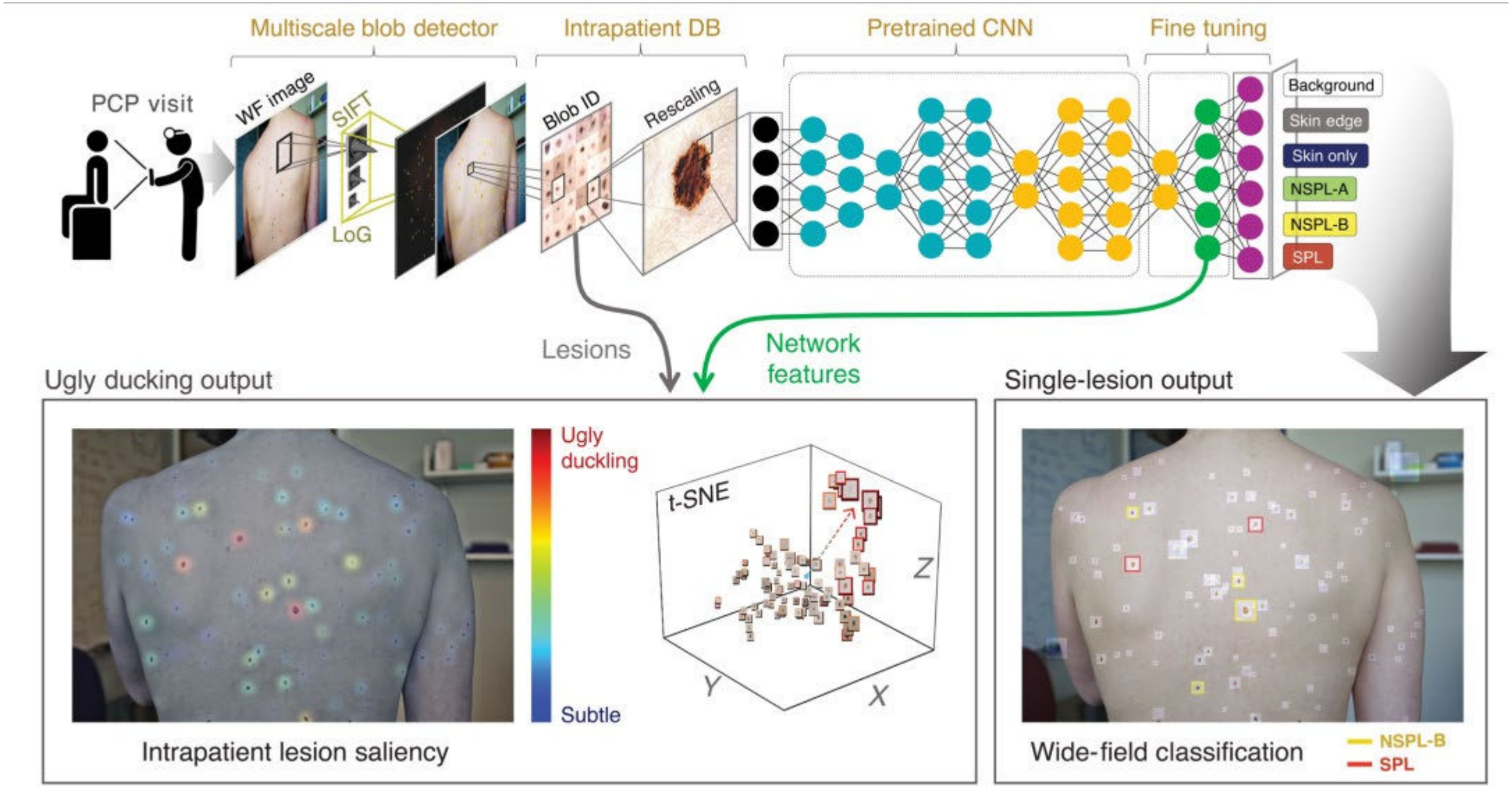
Images courtesy of MetaOptima

Using AI to identify new lesions



Images courtesy of MetaOptima

Evaluating for suspicious lesions and ugly ducklings

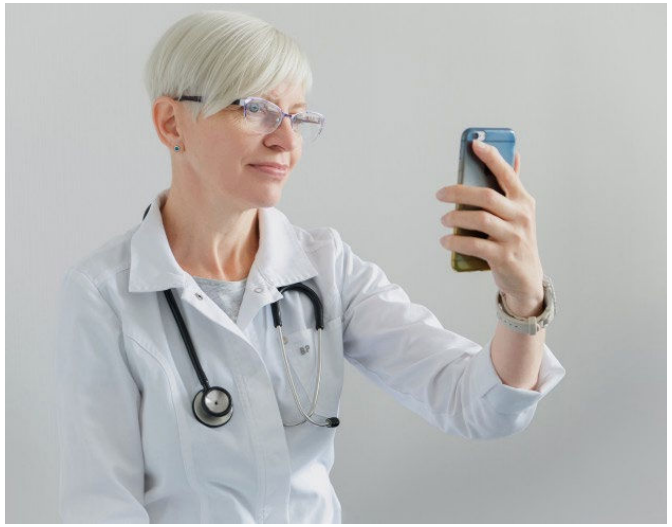


Sci Transl Med. 2021 Feb 17;13(581)

Impact of AI on practice: what may the future hold?

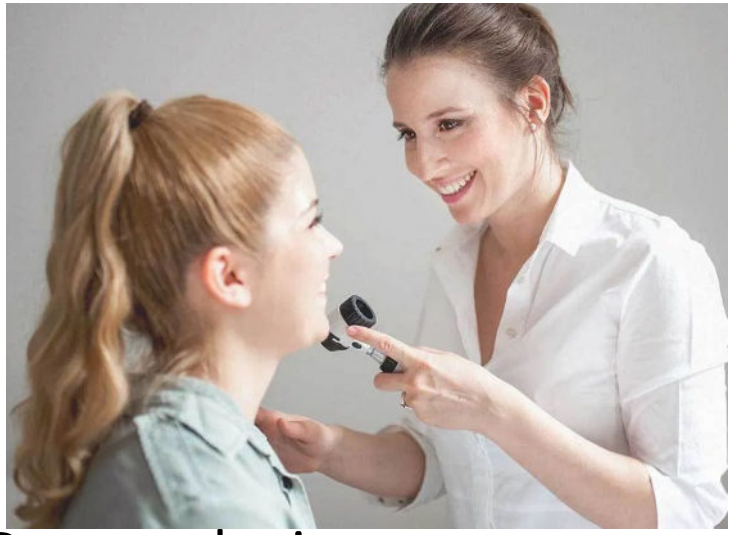
Patients:

- Use smartphone to monitor own nevi (total body, individual)
  - As tracking device or with AI
  - Bias toward high sensitivity → lower specificity
- Visit a pharmacy kiosk to get a “consult” on a suspicious lesion



Non-Dermatologists:

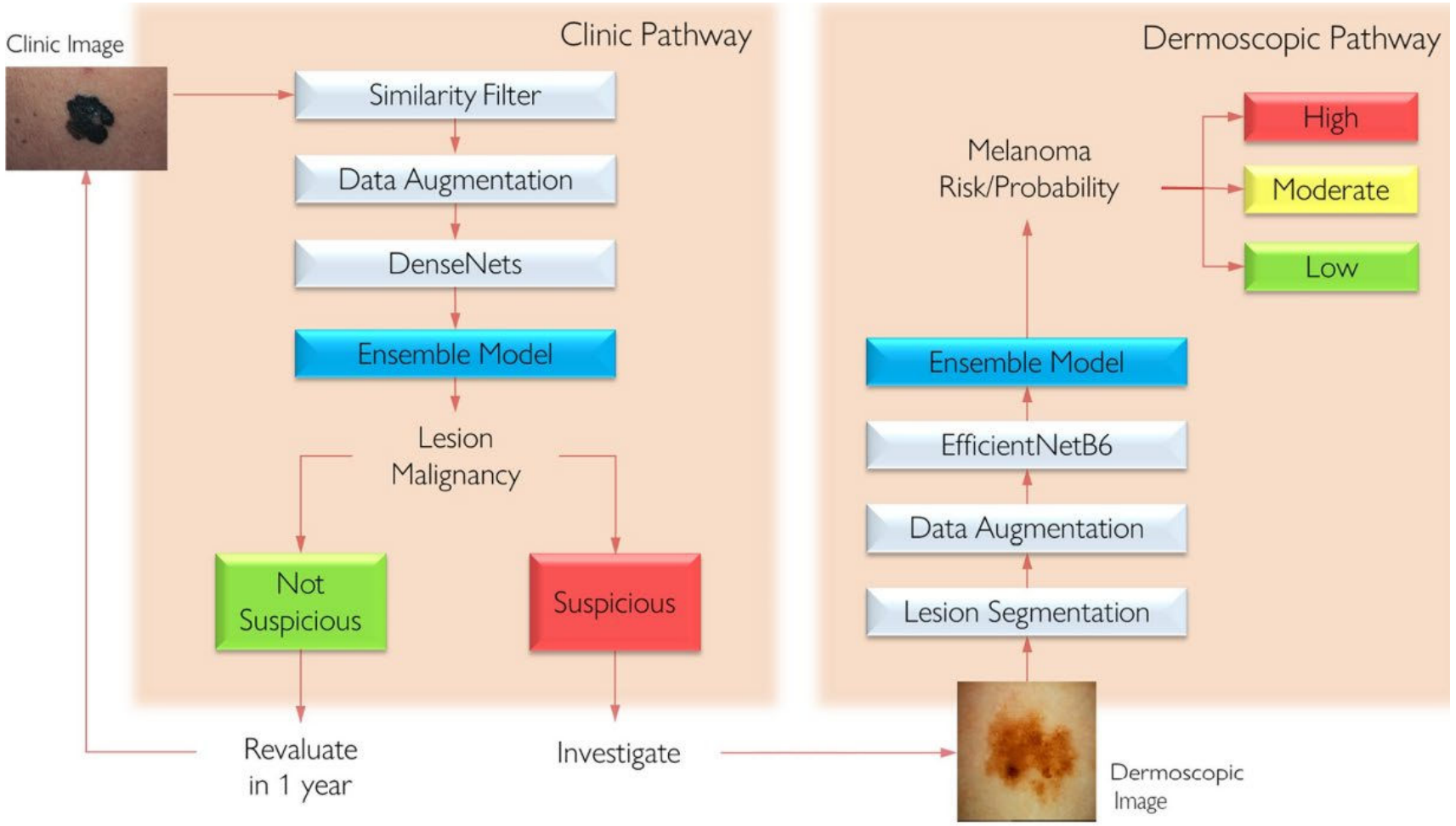
- Tools to triage lesions limit low-yield / benign lesion biopsies and referrals to dermatology



Dermatologists:

- Identify subtle skin cancers
- Streamline the exam, help identify new, changing lesions (vs manual mole mapping evaluation)
- Single lesion evaluation in some cases helpful
- Triage for telemedicine

AI for melanoma diagnosis in primary care



Proposed Triage:

- → Derm fast track
- → Derm usual track
- → Call for changes

Trained using database of available and generated clinical and dermatoscopic images

For clinical photos alone

- Sensitivity : 89%
- Specificity: 85%
- PPV: 58%
- NPV: 97%

CAD system pathways developed to provide outputs for clinical and dermoscopy images to primary care physicians.

PLOS ONE | September 22, 2021

AI to triage lesions for teledermatology

- 100 macroscopic lesions (AK, SK, BCC, SCC, nevus, melanoma (5))– lesions classified by 3 dermatologist and by AI (Triage Snap, Toronto, CA)
- **Overall, dermatologists made more correct diagnoses more often: 78% vs 66% ($p<0.05$)**
- **But, when just considering malignancy, no differences in sensitivity**
- **All melanomas recognized as atypical nevi by AI**
- AI may be helpful in triage of lesions submitted for telemedicine evaluation, but less accurate on macroscopic images vs dermoscopy, worst performance on SK

- J Drugs Dermatol. 2022 Feb 1;21(2):191-194.

Conclusions

- Multiple apps / devices with high (dermatologist-level) sensitivity; variable specificity
- Technologies should be thoughtfully evaluated
  - Important to understand the complexity of the images used for validation
  - Who is the intended user?
  - How are non-melanoma skin cancers classified?
  - Guidelines to standardize evaluation are helpful
  - Important to understand accuracy in skin of color patients
  - Sensitivity must be considered in the context of specificity and user
- Dermatologists miss melanomas and dermpath is not perfect; what sensitivity and specificity are acceptable?