



Geautomatiseerde analyse van huidkanker-app onbetrouwbaar

Y. Chung¹, A.A.J. van der Sande¹, K-P. de Roos², M.W. Bekkenk³, E.R.M. de Haas⁴, N.W.J. Kelleners-Smeets⁵, N.A. Kukutsch⁶

Een wereld zonder apps is niet meer denkbaar. Ook het aantal apps om de huid te checken neemt hand over hand toe. Met deze apps, waarvoor moet worden betaald, kan men een vlek via de telefoon scannen. Een bepaalde rating geeft aan of een vlek een hoog risico heeft of niet. Waar de meeste apps louter de ontwikkeling van bepaalde laesies willen vastleggen, mikt de doelstelling van de SkinVision app (SVA) hoger via de toezegging "to find skin cancer early and get to a doctor at the right time." (www.skinvision.com). Voor de app wordt een algoritme gebruikt die een foto van een laesie als laag, gemiddeld of hoog risico beoordeelt. De consument kan verder ook een Nederlandse dermatoloog (deels tegen extra kosten) inschakelen om de foto virtueel te laten beoordelen. In de door het bedrijf ter onderbouwing aangehaalde studies selecteerde een dermatoloog de te onderzoeken laesies bij patiënten. [1,2] De domeingroep oncologie en het bestuur van de Nederlandse Vereniging voor Dermatologie & Venereologie (NVDV) waren geïnteresseerd in welke mate het oordeel van de app overeenkomt met de diagnose van de dermatoloog. We onderzochten dit onder de bezoekers van de Huidkankerdag 2017, in een prospectieve, multicenter case-controlstudie, in opdracht van de NVDV. Deze setting bood bovendien een mogelijkheid om de app te beoordelen zoals die bedoeld is: bij personen die zelf het plekje kiezen waarover zij ongerust zijn, zonder interferentie van een zorgverlener.

METHODE

Na instemming voor het onderzoek door de medisch ethische commissies, werden bezoekers in vier academische centra - Leiden, Amsterdam, Maastricht en Rotterdam - gevraagd twee plekken aan te wijzen die ze nader onderzocht wilden zien. Door een onderzoeker werd een korte vragenlijst ingevuld en een foto gemaakt van de aangewezen laesie en vervolgens geüpload voor directe beoordeling via de geautomatiseerde analyse van de app. Daarna werd dezelfde laesie beoordeeld door een dermatoloog (of twee dermatologen in consensus), zonder dat de dermatoloog kennis had van de uitslag van de app. De dermatologen beoordeelden benigne laesies als groen, laesies zoals onverdachte atypische moedervlekken als geel en (pre)maligne als rood. Verder noteerde de dermatoloog de diagnose.

RESULTATEN

Van 125 bezoekers werden in totaal 199 door hunzelf aangewezen plekjes nader bekeken (tabel 1). Het eerste probleem was dat het in 90 gevallen (45%) niet lukte een goede foto met de app te maken; in totaal ondernamen we bij iedereen bij wie dit misging, maximaal vijf pogingen.

Tabel 1. Kenmerken deelnemers (n=125).

Geslacht, n (%)	
Man	31 (25)
Vrouw	69 (55)
Onbekend	25 (20)
Leeftijd (gemiddeld ±SD)	51±15
Deelnemers met x aantal laesies, n (%)	
Deelnemers met een laesie	51 (41)
Deelnemers met twee laesies	74 (59)
Totaal aantal laesies, n	199
Locatie van de laesies, n (%)	
Hoofd/hals	39 (20)
Romp	101 (51)
Bovenste extremiteit	32 (16)
Onderste extremiteit	27 (14)

Van de overgebleven 109 laesies was de rating van de app 'groen' (laag risico) voor 34 laesies (31%), 'geel' (gemiddeld risico) voor 30 laesies (28%) en 'rood' (hoog risico) voor 45 laesies (41%). De rating door een dermatoloog was groen voor 91 laesies (84%), geel voor 9 (8%) en rood voor 9 (8%). Vergelijking van de

¹ Arts-onderzoeker, NVDV, Utrecht

² Tot december 2017 bestuursvoorzitter NVDV, nu Dermapark, Uden

³ Afdeling Dermatologie, AMC en VUmc, Amsterdam

⁴ Afdeling Dermatologie, Erasmus MC, Rotterdam

⁵ Afdeling Dermatologie, MUMC, Maastricht

⁶ Afdeling Dermatologie, LUMC, Leiden

rating van de app met de rating door de dermatoloog gaf het volgende te zien: van de rode rating door de app, beoordeelden dermatologen 38 laesies (84%) als groen, 1 (2%) als geel en 12 (13%) als rood. Van de 90 laesies (45%) waarvoor de app geen rating gaf, gaven dermatologen de scores: 73 als groen (81%), 5 als geel (5%) en 6 als rood (13%). De interobserver-overeenkomst tussen app-rating en de rating door dermatologen was zeer laag (gewogen kappa = 0,016; 95%-CI -0,083-0,115; p = 0,737) (tabel 2a en 2b).

Tabel 2A. Rating van de SVA vs rating van de dermatoloog.

		Rating dermatoloog, n			Aantal van de SVA rating, n
		Groen	Geel	Rood	
Rating SVA, n	Groen	27 (79%*)	5 (15%**)	2 (6%***)	34
	Geel	26 (87%*)	3 (10%**)	1 (3%***)	30
	Rood	38 (84%*)	1 (2%**)	6 (13%***)	45
Totale rating dermatoloog, n		91	9	9	109

Percentage van de SVA-rating dat een groene*/gele**/rode*** rating kreeg door de dermatoloog.

Afkorting: SVA, SkinVision Applicatie.

Tabel 2B. Rating van de dermatoloog bij laesies waarbij de SVA geen opname kon maken.

		Rating dermatoloog, n			Aantal van de SVA rating, n
		Groen	Geel	Rood	
SVA geen foto, n		73 (81%*)	5 (6%**)	12 (13%***)	90

Percentage van de laesies waarvan de SVA geen foto kon nemen, waarna de dermatoloog een groene*/gele**/rode*** rating gaf.

Afkorting: SVA, SkinVision Applicatie.

67 procent van de plekjes die door de app als 'rood' werden bestempeld (hoog risico) en 70 procent van de 'gele app-rating' (gemiddeld risico) waren volgens de dermatoloog een gewone moedervlek of een verruca seborrhoeica (tabel 3).

Zoals gezegd kon de app 90 laesies niet analyseren. Deze 90 zijn wel beoordeeld door een dermatoloog, en deze vond in de groep benigne naevi (35), atypische naevi (4), actinische keratosen (8), basaalcelcarcinomen (9), lentigo solaris (7), lentigo maligna (1), fibromen (3), angiomen (3) en nog enkele andere benigne laesies (6).

Tabel 3. Rating van de SVA vs diagnose van de dermatoloog.

		Diagnose dermatoloog, n											Totaal n
		Nauves n.	Atypische naevus	Verruca seborrhoeica	Actinische keratose	BCC	M. Bowen	Lentigo solaris	Lentigo Maligna	Dermato-fibroom	Angioom	Andere benigne laesies	
Rating SVA, n	Groen	20 (59%)	4 (12%)	4 (12%)	0 (0%)	1 (3%)	0 (0%)	3 (9%)	0 (0%)	2 (6%)	0 (0%)	0 (0%)	34
	Geel	19 (63%)	2 (7%)	2 (7%)	0 (0%)	1 (3%)	0 (0%)	3 (10%)	0 (0%)	2 (7%)	1 (3%)	0 (0%)	30
	Rood	15 (33%)	1 (2%)	15 (33%)	1 (2%)	4 (9%)	1 (2%)	2 (4%)	0 (0%)	3 (7%)	0 (0%)	3 (7%)	45
Totale n		54	7	21	1	6	1	8	0	7	1	3	109

Percentage van de SVA-rating dat van de dermatoloog een bepaalde diagnose kreeg.

Afkorting: SVA, SkinVision Applicatie; BCC, basaalcelcarcinoom.



EVALUATIE EN BESCHOUWING

De resultaten laten een zeer hoge foutmarge zien en er is nauwelijks overeenkomst met het oordeel van de dermatoloog. Soortgelijke resultaten zijn ook door andere onderzoeksgroepen gevonden. [3-6] In een recent verschenen review zijn onder andere apps met geautomatiseerde analyse op een rij gezet en men concludeert dat deze niet betrouwbaar en veilig zijn, ondanks het schermen met certificaten. [7] Ook ons onderzoek toont aan dat weinig waarde kan worden gehecht aan de rating van deze app. Een app is vooralsnog alleen te gebruiken om door middel van foto's de ontwikkeling van een huidafwijking bij te houden, zonder daaraan een risico-inschatting te koppelen. In het algemeen constateren wij dat de app bepaalde laesies, maligniteiten en niet-maligniteiten, niet kan analyseren. Eerdere studies rapporteerden percentages tot 26%. [1,4] Mogelijk ligt in onze studie het aandeel hoger omdat de plekken door de gebruiker zelf werden uitgekozen. Hoe gebruikers op deze ontbrekende ratings zouden reageren, blijft ongewis. Zorgelijk is dat een onterecht hoogrisico-oordeel door de app consumenten onnodig ongerust maakt wat weer kan leiden tot hogere en overbodige zorgconsumptie. Het behoeft geen betoog dat het zeer ernstig is als er maligniteiten worden gemist, zoals blijkt uit andere studies die deze app of een soortgelijke anonieme app hebben getest. [4-6] Er zijn ongetwijfeld ontwikkelingen op het gebied van kunstmatige intelligentie die in de toekomst als app beschikbaar komen, maar ook deze technologie moet zich nog in de praktijk bewijzen. [8,9] Het is teleurstellend dat een firma blijkbaar hoge beloften in het vooruitzicht mag stellen, zonder die waar te maken. Het is dringend gewenst dat er regelgeving komt om de consument helderheid te verschaffen over de betrouwbaarheid van de app.

LITERATUUR

1. Maier T, Kulichova D, Schotten K, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *J Eur Acad Dermatol Venereol* 2015;29(4):663-7.
2. Thissen M, Udrea A, Hacking M, von Braunmuehl T, Ruzicka T. Health App for risk assessment of pigmented and nonpigmented skin lesions-A study on sensitivity and specificity in detecting malignancy. *Telemed J E Health* 2017;23(12):948-54.
3. Nabil R, Bergman W, Kukutsch NA. Poor agreement between a mobile phone application for the analysis of skin lesions and the clinical diagnosis of the dermatologist, a pilot study. *Br J Dermatol* 2017;177(2):583-4.
4. Ngoo A, Finnane A, McMeniman E, Tan JM, Janda M, Soyer HP. Efficacy of smartphone applications in high-risk pigmented lesions. *Australas J Dermatol* 2017 Feb 27. doi: 10.1111/ajd.12599. Epub ahead of print.
5. Wolf JA, Ferris LK. Diagnostic inaccuracy of smartphone applications for melanoma detection-reply. *JAMA Dermatol* 2013;149(7):422-6.
6. Dorairaj JJ, Healy GM, McInerney A, Hussey AJ. Validation of a melanoma risk assessment smartphone application. *Dermatol Surg* 2017;43(2):299-302.
7. Rat C, Hild S, Rault Sérandour J, et al. Use of smartphones for early detection of melanoma: systematic review. *J Med Internet Res* 2018;20(4):e135.
8. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
9. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncol* 0:1-7, 2018 doi:10.1093/annonc/mdy166.

SAMENVATTING

Er zijn meerdere apps beschikbaar die claimen huidkanker in een vroeg stadium te kunnen opsporen. Verschillende studies bij patiënten lieten voornamelijk negatieve resultaten zien. De SkinVision app (SVA) biedt de mogelijkheid om via een geautomatiseerde analyse een foto van een laesie te laten beoordelen in laag, gemiddeld of hoog risico voor huidkanker. Wij wilden weten of de rating van de SVA betrouwbaar is in een setting zoals die door de firma is bedoeld: bij personen die zelf hun plek uitkiezen zonder tussenkomst van een zorgverlener. Bezoekers van de Huidkankerdag 2017 konden in vier academische centra maximaal twee voor hun verdachte plekken aanwijzen die door de app werden geanalyseerd. Aansluitend werden dezelfde plekken door een dermatoloog, die geen kennis had van de app-rating, beoordeeld. In de studie werden 125 bezoekers met 199 laesies geïnccludeerd. In 90 gevallen (45%) lukte de analyse door de app niet, waarvan negen basaalcelcarcinomen, vier atypische naevi en een lentigo maligna. Bij 30 (67%) van de hoogrisico- en bij 21 (70%) met een gemiddeldrisicoring door de app stelde de dermatoloog de diagnose benigne naevus of verruca seborrhoea. De interobserverovereenkomst tussen app-rating en de rating door dermatologen was zeer laag (gewogen kappa = 0,016; 95%-CI -0,083-0,115; p = 0,737), vergelijkbaar met de uitkomst van een kansspel. Het is dringend gewenst dat er regelgeving komt voor apps die beloftes doen om de gezondheid te verbeteren zodat de consument en zorgverleners weten welke apps betrouwbaar zijn.

TREFWOORDEN

mobiele telefoon – vroegdiagnostiek – huidkanker – melanoom – basaalcelcarcinoom – geautomatiseerde analyse – app(licatie)

SUMMARY

Several smartphone apps claim to be able to find skin cancer early but studies have shown mainly poor results. However, those study groups consisted of patients whose investigated lesions were mainly selected by a medical professional. The SkinVision app (SVA) automatically analyses pictures of lesions and provides an instant rating as low, medium or high risk for skin cancer. Our aim was to find out whether the rating of the app is reliable in a real-life setting, in which consumers themselves select the lesions without interference of a health professional. Visitors of the National Skin Cancer Day in four university medical centres (in 2017) were asked to select up to two lesions they wanted to be analysed by the app. Then, the same lesions were rated by a dermatologist who was blinded to the rating of the app. One hundred and twentyfive visitors with 199 lesions were included. In 90 cases (45%) the app was not able to perform an analysis (of which nine basal cell carcinoma, four atypical nevi and one lentigo maligna). More than two thirds of the cases with a red or yellow rating by the app were diagnosed as a benign nevus or a seborrheic keratosis by the dermatologist. The interobserver agreement between the ratings of the app and the dermatologist was poor (weighted kappa = 0.016; 95% CI -0.083-0.115; p = 0.737), a value comparable with mere chance. These results stress the urgent need for regulations for apps claiming to have an impact on public health. Up to now a reliable quality mark which could guide consumers and health professionals is still lacking.

KEYWORDS

smartphone application – skin cancer – melanoma – basal cell carcinoma – automated analysis

CORRESPONDENTIEADRES

Nicole Kukutsch

E-mail: n.a.kukutsch@lumc.nl