



# Kwaliteitsbeoordeling van mobiele apps, software en AI: essentieel vóór de inzet in de dermatologie

R.I.F. van der Waal<sup>1</sup>, D. Vellinga<sup>2</sup>, T. Sangers<sup>3</sup>, P.W.O. van Aken<sup>4</sup>

Het gebruik van apps, software en artificial intelligence (AI) in de zorg of in wetenschappelijk onderzoek is niet zonder risico's en is daarom aan voorwaarden gebonden. Als zorgverlener moet men een aantal punten kritisch nalopen voor ingebruikname van een app of software (incl. AI). Hoewel het wenselijk is voor de dermatologische professionals en patiënten meer duidelijkheid te geven over de geboden kwaliteit hiervan, is dit een praktisch onmogelijke taak. Dit artikel geeft een beknopt overzicht van de kritische punten en de momenteel beschikbare hulpbronnen.

Het NVDV-bestuur vroeg de commissie Digitale Dermatologie om informatie richting de leden op te stellen over de omgang met het groeiende aanbod van huidapps, software en artificial intelligence en hoe deze te beoordelen. Dit heeft geleid tot onderstaand geheel.

## INLEIDING

E-health is het gebruik van informatie- en communicatie-technologie (ICT) om gezondheid en zorg te ondersteunen en/of te verbeteren en is een van de belangrijkste ontwikkelingen in de hedendaagse gezondheidszorg. [1] Volgens de Wereldgezondheidsorganisatie (WHO) kan het gebruik van ICT de volksgezondheid effectief, doelmatig en veilig verbeteren. Daarom ondersteunt de WHO sinds 2005 wereldwijd de implementatie van eHealth. Het inzetten van ICT in de zorg gaat echter ook gepaard met risico's. Binnen de gezondheidstechnologie is het gebruikelijk de risico's te onderscheiden naar waar ze zich voordoen. Het Nederlandse kenniscentrum voor landelijke toepassingen van ICT in de zorg (Nictiz) benoemt risico's van e-health op drie fronten: mens, organisatie en technologie. [2] Bij inventarisatie van deze risico's vormt het centraal stellen van de eindgebruiker steeds een solide uitgangspunt.

## POTENTIE EN RISICO'S VAN NIEUWE TECHNOLOGIE

De laatste jaren wordt in de zorg - ook in de dermatologie - steeds meer gebruik gemaakt van mobiele applicaties op



Afbeelding 1.

smartphones, afgekort tot 'apps', en websites als onderdeel van e-health. [3] Met apps kunnen gebruikers eenvoudig gerichte informatie inzien en geïnformeerd worden met notificaties, zogeheten 'pushberichten'. Ook kunnen apps gegevens verzamelen via ingebouwde sensoren en via de netwerkverbinding van het apparaat waarop ze geïnstalleerd zijn. Het wereldwijd gebruik van smartphones biedt eenieder in potentie ook de mogelijkheid om laagdrempelig (zelf)diagnostiek te verrichten via foto's van mogelijk voor maligniteit verdachte huidafwijkingen, waarbij slimme software de huidafwijking beoordeelt. Gezien de sterk stijgende incidentie van huidkanker biedt deze nieuwe technologie in potentie grote (gezondheids)kansen. Het kan leiden tot minder onnodige

<sup>1</sup> Dermatoloog, afdeling Dermatologie, Tergooi MC, Hilversum

<sup>2</sup> Dermatoloog, afdeling Dermatologie, Alrijne ziekenhuis, Leiderdorp

<sup>3</sup> Promovendus, afdeling Dermatologie, Erasmus MC, Rotterdam

<sup>4</sup> Graduate School of Natural Sciences, Artificial Intelligence, Universiteit Utrecht

\* mede namens de overige leden van de NVDV-commissie Digitale Dermatologie: X. Liu, H. Martens, S. Pentinga, M. Seyger, M. Wakkee

consulten bij de huisarts en/of de dermatoloog. Tegelijkertijd kan het bijdragen aan vroegdetectie van huidkanker en daarmee een positief effect hebben op de eerdere en/of beperktere behandeling en de prognose van patiënten met huidkanker. Het gebruik van apps in de zorg of in wetenschappelijk onderzoek is niet zonder risico's en is daarom aan voorwaarden gebonden. [3-5] Ditzelfde geldt voor andere software van onder andere wearables, websites (bijvoorbeeld online medische calculators) en in artificial intelligence (AI). [5,6] De zorgverlener moet een aantal punten kritisch nalopen voordat hij een app of software, waarin ook gebruik gemaakt kan worden van AI, in gebruik neemt, zie kader 1. [3-5]

### APPS EN SOFTWARE ALS MEDISCH HULPMIDDEL

Een app of software (inclusief wearables, websites of AI) - bestemd voor gebruik bij een medisch doel, zoals het diagnosticeren, behandelen, voorkómen, of monitoren van een ziekte, letsel of een beperking - is een medisch hulpmiddel en moet een Conformité Européenne (CE)-markering hebben. [5] De CE-markering geeft aan dat het product voldoet aan de geldende regels binnen de Europese Economische Ruimte nadat de juiste procedures hiertoe zijn doorlopen. [4,7] Het is belangrijk te weten dat CE-markering geen kwaliteitskeurmerk is. [9] Medische hulpmiddelen worden in verschillende risicoklassen ingedeeld naar beoogd gebruik en het risico van het

medisch hulpmiddel voor de eindgebruiker. [4] De regels voor de classificatie van apps en software als medisch hulpmiddel zijn in mei 2021 aangescherpt met strengere veiligheids- en kwaliteitseisen door de inwerkingtreding van de nieuwe Europese verordening voor medische hulpmiddelen, de zogeheten 'medical device regulation' (MDR). [5] Zo vallen apps die gebruik maken van een algoritme om een huidafwijking te beoordelen volgens de MDR bijvoorbeeld minimaal in klasse II, terwijl deze voorheen als klasse I konden worden geregistreerd. Het voordeel van deze nieuwe wetgeving is dat fabrikanten vanwege de herclassificatie (klinisch) wetenschappelijk bewijs moeten kunnen tonen dat een app conform de door de fabrikant gerapporteerde nauwkeurigheid functioneert. Echter is er geen minimale grens vastgesteld waar een app aan moet voldoen, hetgeen, ondanks de aangescherpte wetgeving, nog steeds als een lage reguleringsstandaard kan worden gezien. [6,9,10]

### BEOORDELING VAN KWALITEIT

In het artikel 'Apps in de zorg: wat moet ik weten?' geven de auteurs een overzicht van de relevante wet- en regelgeving en bieden een aantal praktische handvatten voor zorgprofessionals die een app veilig willen gaan inzetten als medisch hulpmiddel of voor medisch-wetenschappelijk onderzoek. [4] Specifiek relevant voor de dermatologie is de komst van

*Kader 1: Kritische punten voor beoordeling van een app of software (incl. AI) [4,7,8]*

#### PRIVACY

De app/software is transparant over het gebruik van data en er is een privacyverklaring in de app die ten minste toestemming vraagt voor: (1) het verzamelen van gegevens, (2) toestemming vraagt tot gegevens op de smartphone (indien van toepassing), (3) het gebruik van (ingevoerde en afgegeven) gegevens, (4) het aanpassen van ingevoerde gegevens, (5) het verwijderen van ingevoerde gegevens, waaronder het account. Wat zijn de mogelijke risico's als de gebruiker de app verkeerd of onveilig gebruikt? Bijvoorbeeld bij verbinding met een onbeveiligd netwerk?

#### BEOOGDE DOELGROEP

Doelgroep van de app/software is duidelijk beschreven.

#### COMPATIBILITEIT

De app/software is compatibel met de meest gebruikte mobiele besturingssystemen (iOS, Android).

#### ADEQUATE CERTIFICERING

Indien een app/software wordt gebruikt als medisch hulpmiddel dient deze in de juiste klasse CE-gecertificeerd te zijn.

#### MEDISCH INHOUDELIJKE INFORMATIE

De app/software dient gebaseerd te zijn op actuele vakkennis en dient zoveel mogelijk in overeenstemming te zijn met geldende richtlijnen en zorgstelsel.

#### TRANSPARANTIE

De app/software is transparant over de gebruikte trainingsdataset alsook hoe het eventueel ingebouwde algoritme tot een classificatie komt.

#### VALIDATIE

De app/software is gevalideerd, onder andere sensitiviteit en specificiteit, door (medisch) onderzoek, bij voorkeur in een vergelijkbare groep als de doelpopulatie, waarbij de resultaten publiekelijk inzichtelijk zijn.

#### UPDATE VAN INHOUD

De app/software ontwikkelaar voert regelmatig controle uit van de inhoud en updatet deze inhoud spoedig op basis van nieuwe inzichten in de (medische) literatuur.

#### KLANTENCONTACT

De app/software-ontwikkelaar is bereikbaar voor vragen of opmerkingen over de inhoud van de app.

#### GEBRUIKSVRIENDELIJKHEID

De app/software is eenvoudig te gebruiken en toegankelijk voor alle leeftijden.

#### BETAALBAARHEID

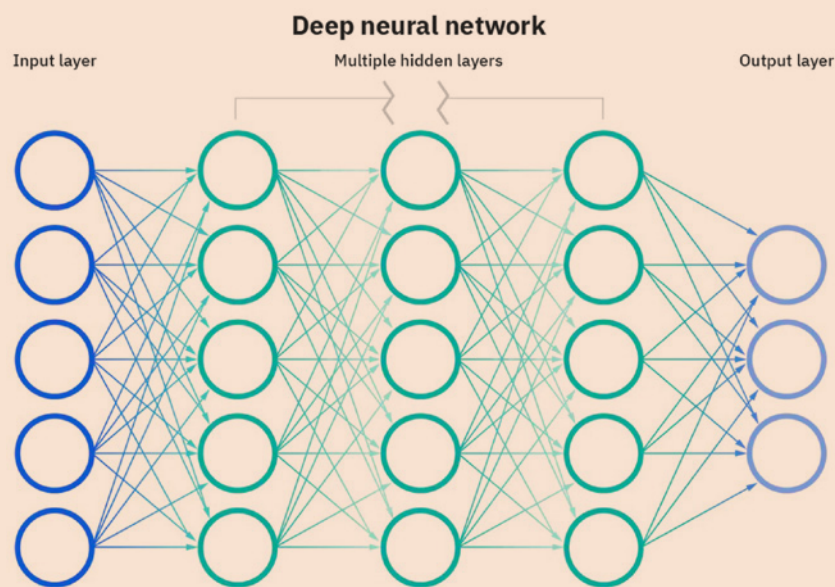
De app/software is voor iedereen betaalbaar.

Met **artificial intelligence (AI)** wordt het vermogen van een computer bedoeld om taken uit te voeren die normaliter door een mens worden uitgevoerd.

**Machine learning** is een vorm van AI waarbij een algoritme getraind wordt om gegevens te leren en op basis daarvan voorspellingen te doen. Machine learning algoritmes worden nauwkeuriger naarmate er meer data wordt gebruikt om het algoritme te trainen. **Deep learning** is onderdeel van machine learning waarin gebruikt wordt gemaakt van **neurale netwerken**, een geavanceerd type algoritme. Deze neurale netwerken bestaan uit knooppunten ('nodes') en met elkaar verbonden zijn in verschillende lagen (afbeelding 2). De naam van deze algoritmes komt voort uit de gelijkenis met neuronen en synapsen uit het menselijke brein. Binnen de dermatologie maken onderzoekers bijvoorbeeld van deep learning gebruik door een neuraal netwerk te trainen op basis van een groot aantal foto's van huidafwijkingen die vooraf gelabeld zijn met een diagnose. Deze vorm

van trainen wordt ook wel *supervised learning* genoemd. Zo kan bijvoorbeeld een algoritme getraind worden om een melanoom met hoge nauwkeurigheid te onderscheiden van een benigne naevus. Helaas is de methode waarop

een uitkomst door deze algoritmes wordt berekend door de complexe opbouw van schakelpunten en lagen zeer beperkt inzichtelijk. Deep learning algoritmes worden daarom ook wel beschouwd als een 'zwarte doos'.



Afbeelding 2.

AI-algoritmes voor het beoordelen van huidafwijkingen (bijvoorbeeld *machine en deep learning* algoritmes, zie kader 2) die sinds kort in apps worden ingebouwd. De beoordeling van de kwaliteit van deze apps brengt extra uitdagingen met zich mee. Het is belangrijk te weten wat de kwaliteit van de trainingsset van een algoritme is, bijvoorbeeld welke gouden standaard (bijvoorbeeld histopathologische of klinische diagnose) er is gebruikt bij het labelen van de diagnose van huidafwijkingen in de trainingsset. Als de dataset waarmee deze algoritmes worden getraind een bias bevat - denk bijvoorbeeld aan een liniaal afgebeeld op een foto naast een melanoom -, dan zal de output van dit algoritme ook deze bias kunnen bevatten (algoritme geeft aan een melanoom te herkennen bij een foto van een naevus met een liniaal). Het is daarom belangrijk dat apps die een beoordeling geven op basis van een algoritme transparant zijn over de gebruikte dataset voor het trainen van dit algoritme. [11] Daarnaast moet zowel de trainingsset als het algoritme adequaat getoetst zijn op de aanwezigheid van bias in deze dataset en de uitkomst van het algoritme. [12] Tot slot is transparantie hoe het algoritme tot een classificatie komt mogelijk nog belangrijker.

Er komt steeds meer aandacht voor *explainable AI*, dat wil zeggen dat de uitkomsten van het algoritme door mensen kunnen worden begrepen. Momenteel wordt voor het

opsporen van huidafwijkingen op foto's vaak deep learning gebruikt. Daarbij is het nog erg lastig om inzicht te krijgen in het beslissingsproces. Maar er zijn ook een aantal eenvoudigere soorten algoritmes waarin de resultaten wel *explainable* zijn. De *explainability* van AI speelt een cruciale rol voor het vertrouwen van zowel de arts/zorgverlener als de patiënt in het gebruik van dit soort apps. [13]

Daarnaast is het van belang te weten hoe representatief die trainingsset is ten opzichte van de populatie waarin de app/AI in de praktijk gebruikt kan worden. Zo is bijvoorbeeld een gevalideerde app of software/AI ontwikkeld op basis van een dataset uit de tweedelijnde geen garantie dat deze ook in de eerstelijnde of nuldelijnde toepasbaar en valide is. Bij dermatologie is extra van belang ook specifiek te kijken naar de in de trainingsset gebruikte type Fitzpatrick huidskleuren, omdat veel datasets die worden gebruikt een groot zwaartepunt hebben in de richting van de lichtere huidskleuren en daarbij niet automatisch voor andere type huidskleuren hun waarde hebben. [14]

Tot slot is het voor de gebruiker essentieel dat de app/software in begrijpelijke taal aangeeft welke diagnoses herkend kunnen worden en wat de nauwkeurigheid is van het algoritme om de betreffende diagnoses te herkennen. Wanneer men een AI-app beoordeelt, is het doel dat men er mee wil bereiken erg belangrijk. Bij focus op preventie ligt de nadruk vooral bij een hoge sensitiviteit, maar ligt de focus op verlagen van

zorglast dan is specificiteit ook belangrijk. Idealiter zouden we voor de dermatologie over een kwaliteitslabel per app of software willen beschikken. Dit kan zowel voor de dermatologische professionals als voor patiënten meer duidelijkheid over de geboden kwaliteit creëren. Praktisch zou dit vergelijkbaar kunnen zijn met bijvoorbeeld de classificatie van energielabels. Vanwege onder meer de hoeveelheid aanbieders, het gebrek aan validatie(tools) en aan transparantie én het dynamische karakter van apps en software door updates is dit voor de NVDV echter een onmogelijke taak. Als alternatief wijzen we erop dat zorgverleners en patiënten gebruik kunnen maken van de 'Leidraad kwalitatieve diagnostische en prognostische toepassingen van AI in de zorg' om de kwaliteit van AI te beoordelen. [15] Daarnaast kunnen zorgverleners en patiënten bij de beoordeling van apps en software gebruikmaken van de KNMG-checklist voor medische apps, het kennisplatform National eHealth Living Lab (NeLL) of de GGD Appstore. [7,15]

### TREFWOORDEN

e-Health – app - artificial intelligence (AI) - machine learning - deep learning

### GEMELDE BELANGENVERSTRENGELING

Geen

### LITERATUUR

- Krijgsman J, Klein Wolterink G. *Ordering in de wereld van eHealth. Whitepaper 12013. Nictiz; 2012.* <https://www.nictiz.nl/wp-content/uploads/2012/08/Whitepaper-Ordering-in-de-wereld-van-eHealth.pdf>
- Ossebaard H, Idzardi K. *Omgaan met de risico's van eHealth. Whitepaper 13012. Nictiz; 2013.* <https://www.nictiz.nl/wp-content/uploads/2013/11/Whitepaper-omgaan-met-risicos-van-eHealth.pdf>
- Lewis TL, Wyatt JC. *mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res. 2014;16:e210.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4180335/>
- Jansen M, et al. *Apps in de zorg: wat moet ik weten? Ned Tijdschr Geneeskd 2020;164:D4956.*
- Artikel 1 Wet medische hulpmiddelen verwijst voor de definitie van medische hulpmiddelen naar de Verordening (EU) 2017/745 (Medical Device Regulation), <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX%3A32017R0745&from=EN>
- Homans J, Tersmette D, van der Wedden H, Dohmen D. *Smartwatch kan waardevol medisch hulpmiddel zijn. Wel certificering nodig en kijk uit voor ongelijkheid van zorg. Medisch Contact 2021;45:26-8.*
- Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst (KNMG), *Medische App Checker. Utrecht: 2016.* <https://www.knmg.nl/advies-richtlijnen/knmg-publicaties/medische-app-checker.htm>
- Harskamp RE, de Vijlder HC, Bekkenk MW. *Smartphone-apps voor zelfdiagnostiek van huidkanker. Meerwaarde en valkuilen. Ned Tijdschr Geneeskd. 2022;166:D6586*
- Ekker A, van Rest B. *Medische apps, is certificeren nodig? Whitepaper 13005. Nictiz;2013.* <https://www.nictiz.nl/wp-content/uploads/2013/05/13005-Whitepaper-medische-apps.pdf>
- Matin RN, Dinnes J. *AI-based smartphone apps for risk assessment of skin cancer need more evaluation and better regulation. Br J Cancer 2021;124:1749-1750.* [doi.org/10.1038/s41416-021-01302-3](https://doi.org/10.1038/s41416-021-01302-3)
- Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. *Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. JAMA Dermatol. 2021;157:1362-1369.* [doi:10.1001/jamadermatol.2021.3129](https://doi.org/10.1001/jamadermatol.2021.3129)
- Daneshjou R, Barata C, Betz-Stablein B, et al. *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. JAMA Dermatol. 2021.* [doi:10.1001/jamadermatol.2021.4915](https://doi.org/10.1001/jamadermatol.2021.4915)
- Shin D. *The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. Int J Human-Computer Studies. 2021;146(83):102551.* [doi:10.1016/j.ijhcs.2020.102551](https://doi.org/10.1016/j.ijhcs.2020.102551)
- Groh M, Harris C, Soenksen, et al. *Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;1820-8.*
- van Smeden M, Moons C, Hooft L, Kant I, van Os H, Chavannes N. *Leidraad voor kwalitatieve diagnostische en prognostische toepassingen van AI in de zorg. Ministerie van Volksgezondheid, Welzijn en Sport. 21 december 2021.* <https://www.datavoorgezondheid.nl/wegwijzer-ai-in-de-zorg/documenten/publicaties/2021/12/17/leidraad-kwaliteit-ai-in-de-zorg>

### BRON AFBEELDINGEN

- Raad voor Europa, <https://www.coe.int/en/web/artificial-intelligence/-/safeguarding-human-rights-in-the-era-of-artificial-intelligence>
- IBM, <https://www.ibm.com/cloud/learn/neural-networks>

### CORRESPONDENTIEADRES

Rutger van der Waal

E-mail: [rvanderwaal@tergooi.nl](mailto:rvanderwaal@tergooi.nl)