



Beeldalgoritmes in de dermatologie

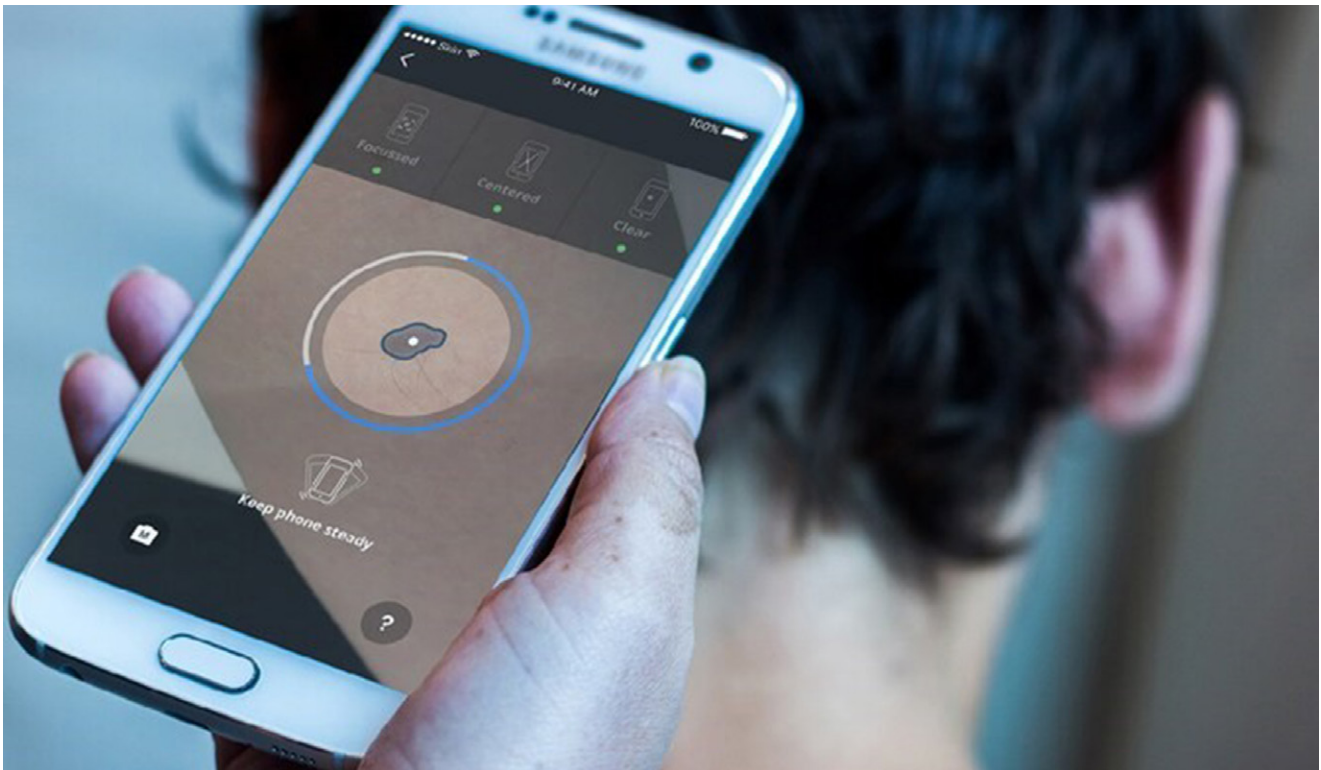
R.C. Beljaards

DISCLAIMER
Deze publicatie is tot stand gekomen **zonder** enige tussenkomst of bemoeienis van artificiële intelligentie (ChatGPT, Bard)

Diagnostiek is effectief gebruik maken van zintuigen. Waar het ene specialisme voornamelijk gebruik maakt van luisteren (denk bijvoorbeeld aan cardiologie en psychiatrie), maakt de dermatologie bij uitstek gebruik van zien. Er is geen andere specialist dan de dermatoloog – de radioloog en de patholoog daargelaten – die via het kijken naar beelden tot een diagnose zal komen. Voorheen waren deze beelden uitsluitend te vinden in leerboeken met dermatologische plaatjes, maar dankzij de digitalisering is het internet de bron geworden waar afbeeldingen van dermatosen te vinden zijn. Dankzij het gemak van de digitale camera in de spreekkamer is het momenteel zeer eenvoudig om dermatosen vast te leggen en de beelden direct gekoppeld aan de patiënt op te slaan in het elektronisch patiëntendossier (EPD). In toenemende mate heeft deze werkwijze voor een deel onze verslaglegging in het EPD van dermatosen via de efflorescentieleer vervangen; het is gemakkelijker, sneller, en reproduceerbaarder voor niet-dermatologen. Het valt zelfs op dat coassistenten minder

bedreven zijn in het beschrijven via het PROVOKE-systeem, mogelijk ligt de tijdsgeest van digitale fotografie daaraan ten grondslag.

Kortom, met zijn allen beschikken dermatologen over grote hoeveelheden digitale beelden van dermatosen. We slaan ze op in het dossier van de patiënt, nemen ze mee naar voordrachten en delen ze met elkaar bij besprekingen. Ook de huisarts heeft allang de weg naar digitale vastlegging van dermatosen gevonden. Dankbaar maken zij gebruik van onze expertise om de door hen opgeslagen beelden te analyseren en te beoordelen. [1] Dit diagnosticum genaamd “teledermatologie” klinkt als bijzondere artificiële intelligentie, maar dat is aanmatigend. Platgeslagen zijn teledermatologie bedrijven niets anders dan intermediairs die foto's rondsturen van de eerste naar de tweede lijn. Vervolgens worden die foto's door ons heel analoog beoordeeld, kortom slechts ‘human intelligence’. Voor het leveren van onze diensten hebben wij (maar ook de huisarts als gebruiker) met deze bedrijven over-



Dermatoloog Centrum Oosterwal, Alkmaar/Heemstede en Sandstep Healthcare, Bosch en Duin | Hoofdredacteur NTvDV

eenkomsten getekend als ware het een toelatingsovereenkomst met een (digitaal) ziekenhuis. De foto's die we mogen beoordelen zijn noch eigendom van de huisarts, noch van de dermatoloog. Het gebruikersrecht van de foto ligt bij de tele-dermatologie instanties.

De volgende logische stap in het beoordelen van dermatologische beelden betreft het creëren van een situatie waarbij de huisarts niet meer aanwezig is; de patiënt wordt consument en is in die situatie zelf in *the lead*. Via een app kan hij/zij een foto uploaden die door een dermatoloog voor een (beperkt) bedrag wordt beoordeeld, de diagnose met behandelplan gaat rechtstreeks naar de consument terug. Deze digitale weg naar de dermatoloog gaat nog steeds via een mens van vlees en bloed, en is dus nog steeds 'human intelligence'.

Van echt een volgend niveau zijn digitale platformen als SkinVision, waarbij niet alleen de insturende huisarts, maar ook de menselijke beoordelaar vervangen is; niet langer velt een dermatoloog de diagnose maar artificiële intelligentie. Een algoritme laat metingen los op het beeld en komt aan de hand daarvan tot een conclusie. In geval van SkinVision betreft het slechts het beantwoorden van pluis/niet pluis als vraagstelling, maar uiteindelijk zal een volgende app zelfs tot een diagnose moeten kunnen komen. Vooralsnog is dit verre toekomstmuziek, want voorlopig kleven alleen al aan de binaire diagnostiek pluis/niet pluis tekortkomingen. [2] Waardoor de positieve uitslagen in de app zelfs nog door een achter de app zittend 'tele'-dermatoloog op de conventionele manier worden gecheckt. Een bijkomstig nadeel is dat apps geen doorlopend zelflerend vermogen kennen; er vindt immers geen terugkoppeling van de uiteindelijke correcte diagnose naar de app om de voorspelling te verifiëren en in de toekomst bij te sturen.

Wat de app als bijvangst realiseert, is een immense database verzamelen van digitale beelden. Net als Captcha ("klik elk vakje met een stoplicht aan") dat we allemaal zonder te morren braaf doen als we een bepaalde website willen bereiken, waardoor we Google helpen om hun beeldherkenning te verbeteren, wordt ook de SkinVision app door de consument onbaatzuchtig gevuld met een veelvoud aan beelden van gepigmenteerde afwijkingen. En de database is hetgeen dit soort platforms waarde geeft. De vraag is wie de juridisch eigenaar is van al die beelden die in deze databases (ongeacht teledermatologie bedrijven of SkinVision) opgeslagen worden? Waarom kunnen we als NVDV niet zelf teledermatologie diensten leveren aan de eerste lijn? En waarom bouwen we daarmee als NVDV niet zelf zo'n database? Met de resultaten van het histologisch onderzoek kunnen we een echt zelflerend algoritme creëren. En wat zou dat algoritme met onze foto's allemaal kunnen doen?

Beeldherkenning is een noodzakelijk onderdeel van de dermatologie. We doen de hele dag in onze spreekkamers niets anders. De beeldherkenningsintelligentie zoals wij die in deze context kennen, maar dan gestuurd door artificiële intelligentie, zijn vaak neurale netwerken. Daarin wordt als het ware een brein nagebootst door kunstmatige neuronen aan elkaar

te verbinden die, afhankelijk van hoeveel signaal ze binnenkrijgen, kunnen kiezen een impuls aan het opeenvolgende neuron door te spelen. Een netwerk met meer neuronen kan getraind worden om complexere problemen op te lossen. Beeldherkenningsalgoritmes maken onderscheid tussen verschillende objecten aan de hand van hun 'features' (karakteristieken/eigenschappen). Een *convolutioneel* neuraal netwerk (CNN) leert uit zichzelf een aantal filters om deze features te detecteren. Bij het evalueren beweegt een filter van links naar rechts en van boven naar onder over het plaatje. Een filter voor het detecteren van een verticale lijn kan bijvoorbeeld kijken of de linker pixels onder het filter donker zijn en de rechter pixels licht. Zo ja, dan geeft dit filter op deze plek een hoge impuls. Op deze manier wordt per filter een *feature map* van impulsen opgebouwd die aangeeft waar in de afbeelding de lijnen en kleurverschillen zitten. Opeenvolgende lagen analyseren in plaats van de originele afbeelding deze feature maps op een steeds hoger niveau, om zo tot een beslissing te komen. Deze beslissing kan een getal zijn tussen 0.0 (zeer pluis) en 1.0 (helemaal niet pluis). De makers van het netwerk kunnen ten slotte een drempelwaarde instellen die bepaalt vanaf welke waarde alarm wordt geslagen.

Het overgrote deel van de beeldherkenning via CNN in de dermatologie is toegespitst op gepigmenteerde afwijkingen. Voor het implementeren van een CNN is geen domein-specifieke kennis meer nodig, maar zijn grote hoeveelheden data nodig voor het trainen.

Het eerste belangwekkende artikel dat verscheen waarin een deep learning convolutioneel neuraal netwerk vergeleken werd met de expertise van een groep dermatologen was meteen schokkend; de groep internationale dermatologen, waaronder 30 experts op het gebied van dermatoscopie, werd verslagen door de prestatie van het CNN. [3] Een nadeel van een CNN, als ze getraind wordt op een niet goed gebalanceerde dataset, is dat zowel de sensitiviteit als de specificiteit enorm kunnen dalen wanneer het algoritme wordt gebruikt om foto's te beoordelen die geen, of onvoldoende, onderdeel zijn van de trainingsset. Een ongebalanceerde dataset bevat niet alleen onvoldoende representatieve foto's van huidafwijkingen die in de praktijk wél zullen worden gefotografeerd maar ontbreekt ook aan bepaalde versturende factoren die in de praktijk aanwezig kunnen zijn, zoals bijvoorbeeld chirurgische markeringen naast een huidafwijking. Een ongebalanceerde dataset kan hierdoor leiden tot een daling van zowel sensitiviteit als specificiteit. Dit probleem kan worden voorkomen door een meer evenwichtige dataset aan beelden aan te bieden. [4]

De beschikbaarheid van evenwichtige data zijn voor dit soort neurale netwerken dus van groot belang. Een neuraal netwerk werkt met behoorlijk veel data (in dit geval digitale beelden van huidziekten). Hoe precies het begrip 'big data' te definiëren is echter moeilijk. Een verzameling van 1000 foto's klinkt al gauw als veel, maar voor big data begint het idealiter pas bij een petabyte (10^{15}) aan foto's (ter vergelijking: het geheugen in een menselijk brein is grofweg 2,5 petabyte). [5] Dit zijn zulke hoeveelheden, dat in de praktijk vaak concessies worden

gedaan en wordt getraind met minder data. Onduidelijk is hoeveel data überhaupt beschikbaar zijn voor beeldalgoritmes in de dermatologie.

Naast kwantiteit om een neurale netwerk te trainen, is verder belangrijk dat de aangeboden set van data divers genoeg is. Tenslotte is in beeldherkenning 'bias' een interessant probleem. Een huidtype boven Fitzpatrick III werkt in beeldherkenning veel minder goed dan een lichte huid. [6] Ook belichting, en andere variaties in de fotografie (zoals de hoek van fotografie) leiden tot een bias.

Waar momenteel nog veel aandacht bestaat voor single lesion deep learning ("is deze afwijking een melanoom of niet?"), ontstaat ook steeds meer interesse in 'ugly duckling algoritmes', met name met betrekking tot het herkennen van afwijkende moedervlekken. Hierbij wordt niet alleen een solitaire afwijking aangeboden aan het CNN, maar een heel huidareaal met meerdere. [7] Apparaten als digitale moedervlek analyse systemen werken volgens dit principe. In deze systemen moet er ook meer aandacht gaan naar patiënt context informatie: de betreffende huidafwijking in relatie met de rest van het huidoppervlak, maar ook klinische informatie zoals door de patiënt zelf opgemerkte veranderingen en subjectieve klachten.

Uiteindelijk zit de beperking van het gebruik van beeldherkenning in de betrouwbaarheid van het algoritme. Zelfs bij een sensitiviteit van 95%, wat als heel krachtig geoordeeld kan worden, zal de implementatie beperkt zijn. Immers, elk gemist melanoom is er één teveel, iets wat door tussenkomst van menselijk gestuurde diagnostiek vanwege de live beoordeling (3D) en verschillende lagen van denken (uitvragen klachten patiënt, afnemen van histologisch onderzoek) ondervangen zou zijn. Daarin wijkt implementatie van een AI-systeem af van het gebruik van menselijke kennis en vaardigheid. We accepteren wel dat een buschauffeur een aanrijding kan veroorzaken, maar van een zelfrijdende Tesla niet. In navolging geldt dat ook voor AI-diagnosesystemen in de zorg. We vinden bij het herkennen van een melanoom een sensitiviteit van 70% door een huisarts aanvaardbaar [8], maar niet voor een huidapp. Bovendien zal een fabrikant omwille van claims dat ook niet acceptabel vinden. De kans dat een algoritme dat geen 100% sensitiviteit bereikt eigenstandig gebruikt gaat worden voor medische diagnostiek is daarom klein. Wel zal ze niet alleen een rol kunnen spelen in het verlangen van de zorgconsument om zelf controle te hebben, maar bovenal ondersteunend zijn aan het diagnostisch handelen van de dermatoloog. Verdere ontwikkeling en verbetering van CNN's heeft de krachtige potentie om uiteindelijk de praktijkvoering van ons vak in hoge mate te veranderen.

We zijn een kwart eeuw verder sinds *Deep Blue* de schaaklegende Garri Kaparov wist te verslaan, maar die was nog regel voor regel met de hand geprogrammeerd. Toen AlphaGo in 2016 go-grootmeester en 18-voudig wereldkampioen Lee Sedol versloeg, had de computer dit voornamelijk aan zichzelf te

danken dankzij een zelflerend vermogen. De afgelopen jaren maakt artificiële intelligentie een exponentiële versnelling door in prestaties. Generatieve AI-modellen als ChatGPT en Midjourney zijn inmiddels op elke mobiele telefoon te downloaden. Critici en wetenschappers waarschuwen voor deze enorme versnelling in lerend vermogen en wijzen op het risico dat artificiële intelligentie de mensheid zal vernietigen. Ook de insteek van deze editie van het tijdschrift ("dermatologen met artificiële intelligentie zullen dermatologen zonder artificiële intelligentie vervangen") klinkt alsof donkere wolken samenpakken boven ons vak, maar we moeten ons als dermatologen geen zorgen maken voor die toekomst. We moeten op tijd meebewegen met de nieuwe ontwikkelingen. En verder: moeten we als gemeenschappelijke dermatologen (lees: NVDV) doorgaan met het onbaatzuchtig leveren van kennis en foto's aan derde partijen? Voorlopig lijkt het voor teledermatologie bedrijven een gemiste kans om foto en diagnose (ze beschikken immers over beide) niet te koppelen om een zelflerend algoritme ten dienste te staan. Als NVDV kunnen we als geen ander vele miljoenen foto's verzamelen, voorzien van een door een dermatoloog geverifieerde diagnose. Het is niet overbodig om zelf deze database van big data te beheren, er zal snel grote behoefte aan zijn.

LITERATUUR

1. Marwaha SS, Fevrier H, Alexeeff S, et al. Comparative effectiveness study of face-to-face and teledermatology workflows for diagnosing skin cancer. *J Am Acad Dermatol*. 81; 1099-106
2. Chung Y, van der Sande AAJ, De Roos K-P, et al. Geautomatiseerde analyse van huidkanker-app onbetrouwbaar. *Ned Tijdschr Dermatol Venereol*. 2018; 28: 11-3
3. 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. *Ann Oncol*. 2018; 29: 1836-42.
4. Pham TC, Luong CM, Hoang DD, Doucet A. AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function. *Sci Rep*. 2021; 11: 17485
5. Wehner MR, Levandoski KA, Kulldorff M, Asgari MM. research techniques made simple: an introduction to use and analysis of big data in dermatology. *J Invest Dermatol*. 2017; 137: 153-8
6. Daneshjou R, Vodrahalli K, Novoa RA, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances*. 2022; 8: 32
7. Soenken LR, Kassir T, Conover ST, Marti-Fuster B, et al. Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Sc Transl Medicine*. 2021; 13: 581
8. Chen SC, Bravata DM, Weil E, et al. A comparison of dermatologist's and primary care physicians' accuracy in diagnosing melanoma. A systemic review. *Arch Dermatol*. 2001;137:1627-34.

CORRESPONDENTIEADRES

Rob Beljaards

E-mail: r.beljaards@nvdv.nl