



WAT AI VOOR DE MEDISCHE
WERELD KAN BETEKENEN

DIGITALE DO



KTERS

Tekst: Hidde Tangerman

Kunstmatige intelligentie kan sommige ziektebeelden beter diagnosticeren dan artsen. Toch heeft deze baanbrekende methode nog de nodige haken en ogen.

Nadat het kunstmatig intelligente softwaresysteem Watson van het techbedrijf IBM in 2011 de beste menselijke deelnemers aan de Amerikaanse spelshow *Jeopardy!* had verslagen, waren de superlatieven niet van de lucht. Watson zou onze nieuwe *computer overlord* worden en ons op alle vlakken van het leven overtreffen met kennis, advies en inzichten waar we zelf nooit op waren gekomen.

IBM zelf juichte het hardst van iedereen. Een dag na die spectaculaire overwinning kondigde het bedrijf aan dat Watson zou worden ingezet in de medische wereld. Als digitale dokter zou het systeem op basis van de wereldwijd verzamelde medische literatuur volmaakte diagnoses kunnen stellen. Zo'n negen jaar later is er van dat optimisme weinig meer over. Diagnoses stellen in de echte wereld blijkt een stuk ingewikkelder dan spelshowvragen beantwoorden. Watsons voorspellingen en adviezen waren vaak onjuist en soms ronduit gevaarlijk. Het systeem heeft nog geen enkel medisch wapenfeit laten zien. Een klassiek geval van hoogmoed voor de val. Toch is kunstmatige intelligentie (AI) bezig aan een indrukwekkende opmars in de medische wereld, zij het met beduidend minder tromgeroffel.

Zelflerende systemen

Die opmars is eigenlijk niet zo gek, want als er één ding is waar AI goed in is, dan is het wel patroonherkenning - en de medische wereld hangt van patronen aan elkaar. Neem het detecteren van kwaadaardige tumoren. Monnikenwerk waarbij medische professionals zoals radiologen en pathologen röntgenfoto's en computerscans uitpluizen op zoek naar afwijkend weefsel. Voor een getraind algoritme zijn deze taken >>>

SHUTTERSTOCK

■ Diagnostiek is een van de moeilijkste onderdelen van het doktersvak. In Nederland is naar schatting 10 tot 15 procent van alle diagnoses incorrect.



■ Van ongeveer de helft van alle medische ingrepen in Nederland is niet bewezen dat ze effectief zijn; 5 tot 10 procent is bewezen nutteloos.

■ Het Amerikaanse kankerinstituut MD Anderson gaf 62 miljoen dollar uit aan een mislukt experiment om IBM's Watson in te zetten voor kankeronderzoek.



■ Nadat het kunstmatig intelligente softwaresysteem Watson de Amerikaanse spelshow *Jeopardy!* won, wilden zijn makers het inzetten om ziektes op te sporen. Dit bleek helaas geen succes.

QUOTE

“Kunstmatige intelligentie heeft vijftig jaar lang beloofd de medische wetenschap radicaal te veranderen, zonder daarin te slagen. Daarom is het belangrijk de nieuwe resultaten niet te overinterpreteren.”

Andrew Beam en Isaac Kohane, onderzoekers aan de Harvard Medical School in Boston, over de recente successen van AI-systemen in de zorg.

» inmiddels een fluitje van een cent. Wetenschappers komen met de ene na de andere studie waarbij hun algoritmes medische problemen detecteren, variërend van botbreuken tot hartritme stoornissen.

Deze opmars van zelflerende systemen in de zorg begon in 2016, toen een groep Amerikaanse en Indiase wetenschappers een algoritme wist te produceren dat diabetische retinopathie (een oogziekte die voorkomt bij diabetici) kon identificeren door foto's van het netvlies te analyseren. Kort erna brachten wetenschappers van de Stanford-universiteit een algoritme naar buiten dat melanomen (huidkankertumoren) kon identificeren aan de hand van smartphonefoto's van stukjes huid. En in 2017 ontwikkelden onderzoekers van het Radboudumc in Nijmegen een zelflerend systeem dat uitzaaiingen van borstkanker succesvol opspoorde door het algoritme scans van lymfeklierweefsel voor te leggen.

Het is een handjevol voorbeelden van een trend die van de schreeuwerige marketingafdeling van IBM is overgeslagen naar wetenschappelijke laboratoria. Het meest opzienbarende van de studies? De algoritmes waren even goed en

vaak zelfs beter dan ervaren medici. Ook mentale stoornissen kunnen inmiddels via AI worden gedetecteerd. Zo ontwikkelden wetenschappers van de Universiteit van Illinois het algoritme DeepMood, dat op basis van de manier waarop mensen het toetsenbord van hun smartphone gebruiken met 90 procent nauwkeurigheid kan bepalen of iemand depressief is. Een ander voorbeeld is de app Companion, die aan de hand van een opgenomen spraakbericht stemmingswisselingen en depressies kan detecteren.

Gratis second opinion

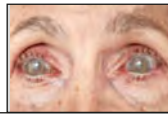
Hoe werkt zo'n zelflerend systeem precies? Enkele van de eerdergenoemde computerwetenschappers van het Radboudumc ontwikkelden onlangs een zelflerend systeem dat prostaatkanker detecteert, waarbij het tevens aangeeft hoe agressief de kanker is. “Die agressiviteit is belangrijk”, zegt Wouter Bulten, die het onderzoek leidde. “Als het bijvoorbeeld om een langzaam groeiende kanker gaat, kan een arts ervoor kiezen om af te wachten of minder invasief te behandelen.”

AI-systemen worden getraind met datasets, in het geval van Bultens

onderzoek ruim vijfduizend scans van prostaatweefsels mét en zonder kanker. Aan elk stukje weefsel had een patholoog (een medisch specialist die weefsels analyseert) een zogeheten Gleason-score toegekend, die de agressiviteit van de tumor weergeeft. Door de data aan het algoritme voor te leggen, leert de software patronen te herkennen die bij de verschillende vormen van agressiviteit horen.

Om te beginnen trainde Bulten het algoritme met een deel van de volledige dataset. Vervolgens testte hij het algoritme met de rest van de data, waarbij hij de fouten die het systeem maakte handmatig corrigeerde. Daarna haalde hij de gecorrigeerde dataset als extra training nogmaals door het systeem.

■ AI-systemen die huidkanker opsporen, zijn getraind op datasets met voornamelijk blanke mensen, waardoor ze mogelijk minder effectief zijn bij zwarte mensen.



■ In 2016 ontwikkelde de 16-jarige Kavya Kopparapu een algoritme dat met behulp van een 3D-geprinte lens de oogziekte retinopathie kan vaststellen.

■ Amerikaanse computerwetenschappers ontwikkelden een algoritme dat op basis van Instagramfoto's depressies met 70 procent accuraatheid kon voorspellen.

■ Samen met internetgigant Google ontwikkelden Amerikaanse en Indiase wetenschappers een kunstmatig intelligent systeem dat de oogziekte diabetische retinopathie detecteert. Op deze foto zie je het systeem in actie.



Toen hij het algoritme vervolgens aan een geheel nieuwe dataset blootstelde, bleek het systeem de analyse beter te doen dan tweederde van de menselijke pathologen.

In een vervolgonderzoek ontdekte de onderzoeker dat de beste resultaten worden behaald wanneer de arts en het algoritme samenwerken. Bulten: “Logisch, want mensen maken andere fouten dan computers.” Zo kunnen computers onscherpe foto’s of kapotte weefsels onterecht als tumoren aanmerken. “Zulke vergissingen zijn voor een mens juist heel simpel te herkennen.”

Volgens de Nijmeegse wetenschapper moeten zelflerende systemen worden gezien als een gratis second opinion en zeker niet als vervanging van de arts. AI kunnen ze op plekken waar weinig specialistische kennis aanwezig is wel een belangrijke rol spelen. “In delen van Afrika heb je één patholoog per miljoen mensen. Daar kan AI helpen om de aanwezige kennis naar het niveau van een expert te brengen.”

Ook Bultens collega Geert Litjens, biomedisch ingenieur aan het Radboudumc, denkt niet dat artsen en psychiaters bang hoeven te zijn voor hun baan. “Wat het beoordelen van prostaatkan-

ker betreft doen we het beter dan een arts, maar die voert wel vijfduizend van dat soort taken uit. We kunnen misschien één vijfduizendste van de arts vervangen, maar bij de andere 4999 taken zit hij nog altijd zelf aan de knoppen.” Ziehier zowel de kracht als de zwakte van de medische algoritmes. Op hun eigen afgebakende gebiedje zijn ze superieur, daarbuiten zijn ze zo goed als waardeloos.

Je kunt natuurlijk vijfduizend verschillende algoritmes bouwen die allemaal één medisch deelgebied bestrijken, maar dat is volgens de meeste AI-wetenschappers onbegonnen werk. In plaats daarvan zijn ze naarstig op zoek naar één overkoepelend systeem dat de hele wereld kan begrijpen en interpreteren, ook wel *general AI* genoemd, de heilige graal van het computeronderzoek. “De halve wereld is ermee bezig,” zegt Litjens, “maar we zijn er nog lang niet.”

Afwijkende data

Toch is er ook kritiek op het introduceren van kunstmatige intelligentie in de zorg. Een van de meest gehoorde bezwaren is dat de datasets waarmee algoritmes getraind worden last kunnen hebben van selectie-bias. Dat wil zeggen dat de trainingsdata geen gebalanceerde afspiegeling vormen van de werkelijkheid. Dat kan leiden tot blinde vlekken, verkeerde conclusies of ronduit pijnlijke fouten. Zo zette een gezichtsherkenningialgoritme van Google in 2015 onder een selfie van een stel zwarte mensen de term ‘gorilla’s’, omdat zwarte gezichten in de trainingsdata ondervertegenwoordigd waren.

Ook in de medische wereld heeft



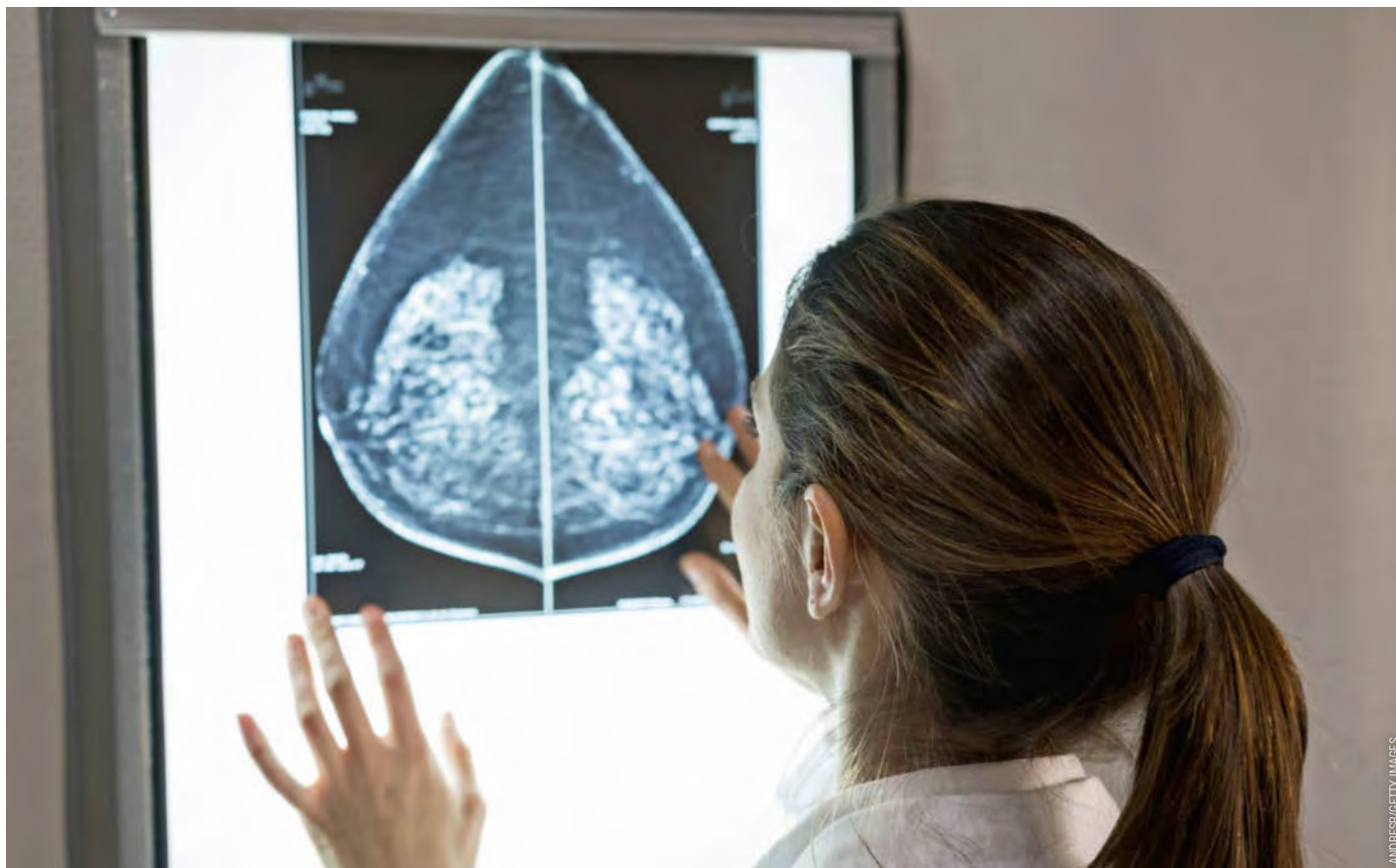
▲ Als gebruikers van de app Companion een dagboek en spraakberichten bijhouden, kan het ingebouwde AI-systeem stemmingswisselingen en depressies bij hen constateren.

■ Er bestaan algoritmes die zelfmoordrisico's inschatten aan de hand van demografische, medische en gedragsdata, zoals hoe vaak iemand lacht, zucht of zich boos uitdrukt.



■ Onderzoekers van het UMC Maastricht hebben een algoritme ontwikkeld dat zwangerschapsrisico's bij aanstaande moeders voorspelt, zoals zuurstoftekort bij de baby.

■ Amsterdamse en Eindhovense computerwetenschappers maakten een algoritme dat beginnende slokdarmkanker negen van de tien keer juist voorspelde.



ANDRESAGRETTY IMAGES

» selectie-bias tot nepnieuws geleid. Het bericht dat matig alcohol drinken beter zou zijn dan helemaal geen alcohol drinken bijvoorbeeld. Dat was gebaseerd op data die alcoholgebruik en ziekenhuisopnames met elkaar vergeleken (niet met behulp van AI overigens), waarbij geen rekening was gehouden met het feit dat sommige geheelonthouders voormalige alcoholverslaafden waren. Ze dronken weliswaar niet meer, maar hadden wel veel ziekenhuisopnames achter hun naam staan. Toen die foute aanname eruit werd gefilterd, bleek dat matig drinken gezonder zou zijn een wassen neus. Zelfs met gevarieerde en representatieve

datasets is het probleem nog niet altijd opgelost. Medische data zien er bijvoorbeeld in elk ziekenhuis weer anders uit. Bulten: “In het ene lab is de temperatuur van het weefsel dat door het algoritme wordt bestudeerd net wat hoger of gebruiken ze een andere scanner. Dat kan invloed hebben op de beelden, waardoor de data kunnen afwijken van de trainingsset waarmee het algoritme is ontwikkeld. Dan wordt de voorspelling ook minder goed.” De oplossing is data van verschillende ziekenhuizen gebruiken, zegt Bulten. Maar dat is vanwege privacybezwaren, verschillen in ICT-systemen en andere obstakels makkelijker gezegd dan gedaan.

▲ Het doorspitten van weefsel om (borst)kanker op te sporen, is een lastig karwei voor artsen. Kunstmatige intelligentie kan ze daar wellicht een handje bij helpen.

Een ander punt van kritiek is dat algoritmen net een *black box* zijn. Omdat de systemen zelflerend zijn, weten zelfs de makers ervan niet hoe hun AI tot bepaalde conclusies komt. In de medische wereld, waar beslissingen op leven en dood worden genomen, is dat gebrek aan transparantie onwenselijk. Per slot van rekening wil een patiënt graag horen hoe een diagnose tot stand is gekomen en waarom een bepaald behandelplan wordt voorgesteld. Toch vindt Litjens die kritiek onterecht. “Artsen kunnen zelf ook niet altijd uitleggen hoe ze tot een beslissing of diagnose komen. En van veel geneesmiddelen die mensen dagelijks slikken, zoals paracetamol, weten we ook niet precies hoe ze werken – alleen dát ze werken. Wanneer AI-systemen het aantoonbaar beter doen dan de mens is de vraag of ze al dan niet transparant zijn niet meer de belangrijkste.”

Minder menskracht?

Volgens Maartje Schermer, hoogleraar filosofie van de geneeskunde aan het Erasmus MC in Rotterdam, loont het juist ook de moeite om te kijken waarin de menselijke dokter beter is dan zijn digitale collega. “Artsen weten wanneer

Blufgids

Algoritme: serie instructies waarmee een computer berekeningen uitvoert om patronen te kunnen herkennen in grote hoeveelheden data.

Zelflerend systeem: kunstmatig intelligente software die op basis van voorgeprogrammeerde kennis gegevens analyseert, verrijkt en adviezen kan geven.

Selectie-bias: een vertekening in de uitslag van

een algoritme veroorzaakt door onevenwichtige trainingsdata.

ROC-curve: grafiek die de kwaliteit van de voorspellende waarde van een algoritme meet.

Waarschuwingsmoeheid: het door een arts gedachteloos wegklikken van computergegenereerde pop-ups op zijn computerscherm.

■ Uit onderzoek is gebleken dat vooral minder ervaren artsen en radiologen steun ervaren van zelflerende systemen, omdat de AI vooral simpele gevallen correct beoordeelt.



■ Kunstmatige intelligentie wordt ook al ingezet voor medicijnonderzoek. De Britse start-up Exscientia bracht het eerste door AI ontdekte medicijn op de markt.

■ AI-bedrijven zijn bezig met het speuren naar een vaccin tegen het coronavirus, maar experts verwachten dat dit nog minstens anderhalf jaar gaat duren.

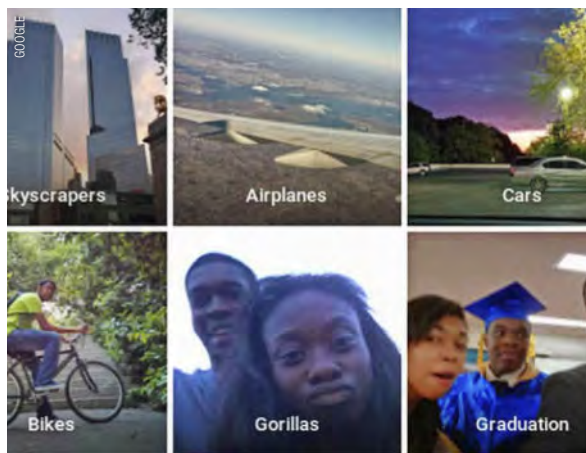


■ Artsen hoeven volgens veel onderzoekers niet bang te zijn dat ze hun baan verliezen door kunstmatige intelligentie. Want dokters kunnen bepalen wat goed is voor een individuele patiënt. AI-systemen hebben deze vaardigheid niet.

een patiënt afwijkt van het gemiddelde, zowel op medisch als sociaal vlak. Een behandeling kan bijvoorbeeld minder wenselijk zijn omdat de patiënt thuis kleine kinderen heeft en niet elke dag naar het ziekenhuis kan komen. Dat een arts van een AI-advies af kan kijken omdat hij of zij ziet wat bij een unieke patiënt past, is heel belangrijk. Computers zijn daar minder goed in.”

Wat artsen ook beter kunnen dan computers is een goed gesprek voeren, empathie tonen en geruststellen. De Amerikaanse cardioloog Eric Topol betoogt in zijn boek *Deep medicine* dat het inzetten van AI er juist voor kan zorgen dat er meer tijd is voor die menselijke kant. Zeker geen overbodige luxe als je bedenkt dat elke extra minuut die een huisarts met een patiënt doorbrengt de kans op een terugkeerbezoek met 8 procent verlaagt, zo toonde een Amerikaans onderzoek uit 2018 aan. Inmiddels hebben bijna duizend studies het positieve verband tussen empathie en gezondheid bewezen.

Ook voor het medisch personeel zelf zal dat een grote verbetering zijn. In de VS krijgt meer dan de helft van de artsen een burn-out (in Nederland is dit 15 procent) en heeft een kwart depressieve



▲ Kunstmatig intelligente systemen zitten er soms faliekant naast. Een gezichtsherkenning algoritme van Google zag twee donkere mensen voor gorilla's aan. Oorzaak: zwarte gezichten waren ondervertegenwoordigd in de trainingsdata.

klachten (Nederlandse cijfers ontbreken), volgens velen het directe gevolg van de enorme stress en tijdsdruk. Toch is het voor Schermer geen uitgemakte zaak dat AI de werkdruk zal verlagen. “Het kan net zo goed de andere kant opgaan, want AI maakt alles efficiënter. Je zou ook kunnen zeggen: de AI is zo goed dat we die dokter ertussenuit halen. Dan gaat dat menselijke contact alsnog verloren.” Maar dat zelflerende systemen hun waarde hebben bewezen, staat volgens Schermer als een paal boven water. In sommige gevallen vraagt zij zich dan ook af of het ethisch verantwoord is om de systemen niet te gebruiken. “Als een computersysteem röntgenfoto's écht beter beoordeelt dan een radioloog, dan is het raar als we daar mensen op zouden zetten.” ■



Hidde Tangerman is freelance-journalist. Voor dit artikel raadpleegde hij onder andere de volgende bronnen: Wouter Bulten e.a.: *Automated deep-learning system for Gleason grading of prostate cancer using biopsies. A diagnostic study*, The Lancet Oncology (1 februari 2020) | Eric Topol: *Deep medicine. How artificial intelligence can make healthcare more human again*, Basic Books (2019).

Ga voor meer informatie naar www.kijkmagazine.nl/artikel/digitale-diagnoses