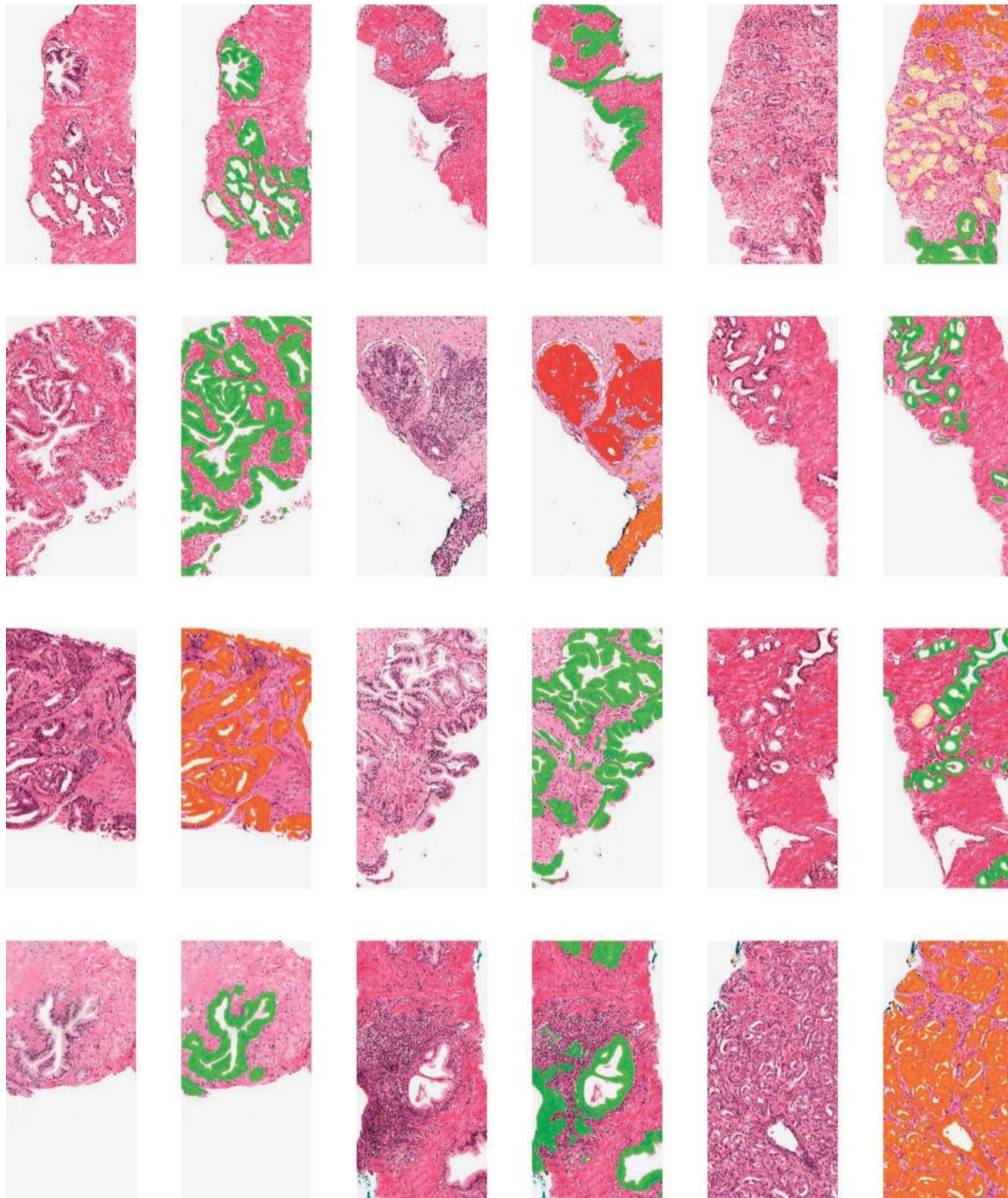


Algoritmes om tumoren op te sporen zijn soms al beter dan de arts.



6/7

Zoekmachine voor kanker

BEELD RABBOUDUMC

2/3

Waterleiding
Schade en angst:
vijf vragen over
lood

5

Karel Knip
Het nut van
koorts meten
op afstand

W
wetenschap

10/11

Sobibor
Het fotoalbum
van een
kampbeul

KUNSTMATIGE INTELLIGENTIE

De software die slimmer is dan de dokter

Zelflerende algoritmes om kanker op te sporen zijn inmiddels even goed als de dokter - of zelfs beter. Willen we dat een computer de diagnoses stelt?

Door onze redacteur **Niki Korteweg**

Kijk, hier liggen de cellen duidelijk rommeliger bij elkaar dan in het gezonde borstweefsel eromheen. En als we inzoomen" - hoogleraar borstpathologie Jelle Wesseling draait de microscoop op een vergroting van 400 keer - „dan zie je dat de celkernen in dit gebied opgeblazen zijn. Er zijn ook meer celdelingen." Behendig schuift hij een ander glaasje met een plakje van een borstbiopt onder de microscoop. Ze zijn naar hem in het Antoni van Leeuwenhoek ziekenhuis opgestuurd om door hem te worden beoordeeld. Zo gaan in Nederland dagelijks honderden glaasjes met verdacht weefsel op de post naar experts. Ze moeten bepalen of een vrouw borstkanker heeft, van welk type die is, en hoe kwaadaardig.

„De twee extremen, zeer agressieve of juist traag groeiende kanker, zijn makkelijk te onderscheiden. Maar daartussen zit een grijs gebied", zegt Wesseling. Elk glaasje wordt daarom bij twielfel door twee pathologen beoordeeld, want ieder kijkt toch net weer anders. De onderlinge scores lopen behoorlijk uiteen, wees onderzoek van Wesselings groep onder negen pathologen uit. En zelfs dezelfde patholoog scoort een beeld de ene dag weer anders dan de andere dag.

Hoe handig zou het zijn om een computer met een slim algoritme in één keer het hele beeld te laten bekijken en de kankerkenmerken te laten scoren? Die werkt snel, raakt nooit vermoeid, beoordeelt altijd volgens dezelfde criteria, en werkt met digitale beelden zodat er geen enveloppen meer heen en weer hoeven.

Enorm handig, dat realiseerden techreuzen als Facebook, Google en Microsoft zich een paar jaar geleden ook, net als talloze kleinere bedrijven. De ontwikkeling van zelflerende kunstmatige intelligentie voor medische toepassingen neemt een hoge vlucht. Slimme algoritmen kunnen al verschillende typen kanker diagnosticeren - en ze zijn minstens even goed als de dokters zelf.

In de eerste weken van dit jaar verschenen kort

na elkaar drie publicaties in hoog aangeschreven wetenschappelijke tijdschriften. Onderzoekers van Google Health en Google Deep Mind presenteerden samen met Britse en Amerikaanse wetenschappers in *Nature* een algoritme dat borstkanker beter op röntgenfoto's kon detecteren dan zes radiologen. Een groep Amerikaanse onderzoekers publiceerde in *Nature Medicine* over hun systeem dat binnen 150 seconden tien verschillende typen hersentumoren van elkaar kon onderscheiden - niet op röntgenbeelden maar op digitale opnamen van stukjes weefsel, nog tijdens een hersenoperatie. En een Nijmeegse groep onderzoekers kwam in *Lancet Oncology* met een algoritme dat niet alleen kan herkennen of er sprake is van prostaatkanker in weefselplakjes, maar ook hoe agressief de tumor is. Zijn die medische artificial intelligence (AI) systemen al klaar voor gebruik? Hoe lang gaat dat nog duren? En zitten pathologen en radiologen er wel echt op te wachten?

Tienduizenden mammogrammen

Twee grote ontwikkelingen maken dat AI nu zo goed wordt in kanker ontdekken. De eerste is dat dankzij bevolkingsonderzoeken een schat aan digitale gegevens beschikbaar is. Daarmee kan een zelflerend algoritme gevoed en getest worden. De tweede is dat sinds een grote doorbraak in 2012 zelflerende (*deep learning*) algoritmes in staat zijn om beelden heel nauwkeurig te analyseren. In dat jaar presenteerde Google bijvoorbeeld een algoritme dat, na het doorploegen van tien miljoen kattenfilmpjes op YouTube, herkende welke dieren wij met het woord 'kat' bedoelen. In de jaren daarop boekten wetenschappers spectaculaire resultaten met AI voor medische toepassingen.

Vooraf voor het beoordelen van röntgenfoto's bij borstkanker is AI al ver ontwikkeld. Het algoritme waarover Google in januari publiceerde, werd getraind met een indrukwekkende hoeveelheid gegevens: de mammogrammen plus de

bijbehorende diagnoses, en informatie over het verloop van de kanker, van bijna 77.000 vrouwen uit een Britse database. Ze testten de AI op de gegevens van bijna 26.000 Britse vrouwen die in het verleden door twee radiologen waren beoordeeld, en die van ruim 3.000 vrouwen uit de VS die door één radioloog was beoordeeld. Het algoritme bleek in beide groepen iets beter te voorspellen of er sprake was van (door een biopt bevestigde) kanker dan de eerste - soms dus de enige - radioloog, en even goed als er twee waren geraadpleegd.

Het algoritme gaf met de Amerikaanse gegevens 5,7 procent minder vals-positieve diagnoses - waarbij iets onterecht als kanker wordt bestempeld, en 9,4 procent minder vals-negatieve diagnoses, waarbij een tumor over het hoofd wordt gezien. Met de Britse gegevens was dit bescheidener: respectievelijk 1,2 procent en 2,7 procent.

Als ultieme test legden de onderzoekers 500 door het systeem beoordeelde mammogrammen voor aan zes Amerikaanse radiologen. Ook in deze test voorspelde het algoritme gemiddeld beter of er sprake was van borstkanker dan de radiologen - maar het miste ook wel eens de juiste diagnose.

„Mooi onderzoek, en de uitkomst verbaast me niet", zegt natuurkundige Bram van Ginneken, hoogleraar medische beeldanalyse aan het Radboudumc in Nijmegen. Hij ziet veel van dit soort studies de laatste tijd. „De software doet het tegenwoordig gewoon beter dan de gemiddelde radioloog." Het Nijmeegse bedrijf ScreenPoint Medical, voortgekomen uit zijn onderzoeksgroep, ontwikkelde vergelijkbare software die, in tegenstelling tot die van Google, al toestemming heeft voor gebruik in de kliniek in Amerika en Europa. „Onze software was beter in het diagnosticeren dan 65 van 101 radiologen. Ik verwacht dat binnenkort AI beter zal zijn dan 90 procent van de radiologen."

Aartsconservatieve medici

De goed scorende radiologen waren over het algemeen ook degenen met de meeste jaren ervaring. „Met dit systeem heb je er dus in feite een ervaren radioloog erbij", zegt Van Ginneken. „Als vrouwen dit zouden weten, zouden ze eisen dat hun mammogram door de software wordt beoordeeld. Je zou bijna zeggen dat het ethisch niet verantwoord is om dat niet te doen. Maar de medische wereld is aartsconservatief", zegt Van Ginneken geïrriteerd.

Want dergelijke software wordt in Nederland en andere Europese landen nog weinig gebruikt. „Er zijn nu tien gecertificeerde AI producten beschikbaar voor het screenen van borstfoto's in de klinische praktijk", zegt Van Ginneken. „In de VS wordt 90 procent van de mammogrammen al gelezen met software. Sinds 2001 krijgen artsen daar een vergoeding voor." Hij vindt het belangrijk dat die er ook in Europa komt, zodat bedrijven de software kunnen ontwikkelen, verkopen en onderhouden.

Radioloog Etta Pisano van de Amerikaanse vereniging voor radiologie, houdt een slag om de arm. De computer-aided detection die in de jaren 1990 werd geïntroduceerd in de VS was veelbelovend in studies, maar bleek in de prak-

tijk tekort te schieten, schrijft ze in een begeleidend commentaar in *Nature*. „Die software was in die tijd nog niet gevoelig genoeg, dat is niet meer te vergelijken met die van vandaag", zegt Van Ginneken.

Oppassen voor loos alarm

Zijn er in Nederland al plannen om bij het bevolkingsonderzoek voor het opsporen van vroege borstkanker AI in te zetten? „Wij kijken zeker naar deze ontwikkelingen, ze zijn veelbelovend. Maar we zoeken nog naar de beste manier om het in te zetten", zegt epidemioloog Mireille Broeders. Aan het Radboudumc in Nijmegen onderzoekt ze de voor- en nadelen van de landelijke borstkankerscreening. „Er is een belangrijk verschil met de VS. Daar is geen centraal geregeld screening, een mammogram is daar dus bedoeld om een diagnose te stellen. Aan onze landelijke screening doen 1 miljoen vrouwen per jaar mee. Het overgrote deel heeft geen borstkanker, die wil je ook niet ongerust maken met een verwijzing die later loos alarm blijkt. We veranderen dus niet lichtzinnig het protocol, we willen die screening zo zorgvuldig mogelijk doen. Nu wordt van elke 1.000 vrouwen er 24 doorverwezen, waarvan er 7 daadwerkelijk borstkanker hebben. We willen er niet meer nodeloos alarmeren, liever juist minder."

„Het zijn veel algoritmes, zegt ze, „maar elk weer net voor een ander kanker stadium of een andere groep vrouwen. Die zijn dan nog niet geschikt voor ons screeningsprogramma." Hoe de software vervolgens ingezet zou moeten worden is ook nog niet duidelijk: als vervanging van de tweede radioloog? Naast het oordeel van de twee radiologen? Of mag de computer bepalen naar welke beelden niemand hoeft te kijken? „Zouden vrouwen dat acceptabel vinden?" vraagt Broeders zich hardop af. „En wie is verantwoordelijk als achteraf blijkt dat de verkeerde diagnose is gesteld? De arts? Of de fabrikant van de software? Daarover moeten we het eerst eens worden."

Terwijl de radiologie met deze vraagstukken worstelt, is AI voor het screenen van microscopische beelden nu ook in opkomst. Het systeem waarmee Amerikaanse onderzoekers tien verschillende hersentumoren kunnen onderscheiden, werd getraind met een ontzagwekkende 2,5 miljoen beelden. Dat waren geen microscopische plakjes maar opnames van stukjes weefsel die met een nieuwe techniek (*stimulated Raman histology*) gedetailleerd in beeld waren gebracht. De AI kon zo tijdens een operatie binnen 150 seconden voorspellen wat de diagnose van de tumor van 278 patiënten was, die een patholoog een half uur later gaf op basis van de gebruikelijke beoordeling. Ook dit systeem deed niet onder voor de menselijke experts: het had 94,6 procent goed, tegen 93,9 procent door de pathologen.

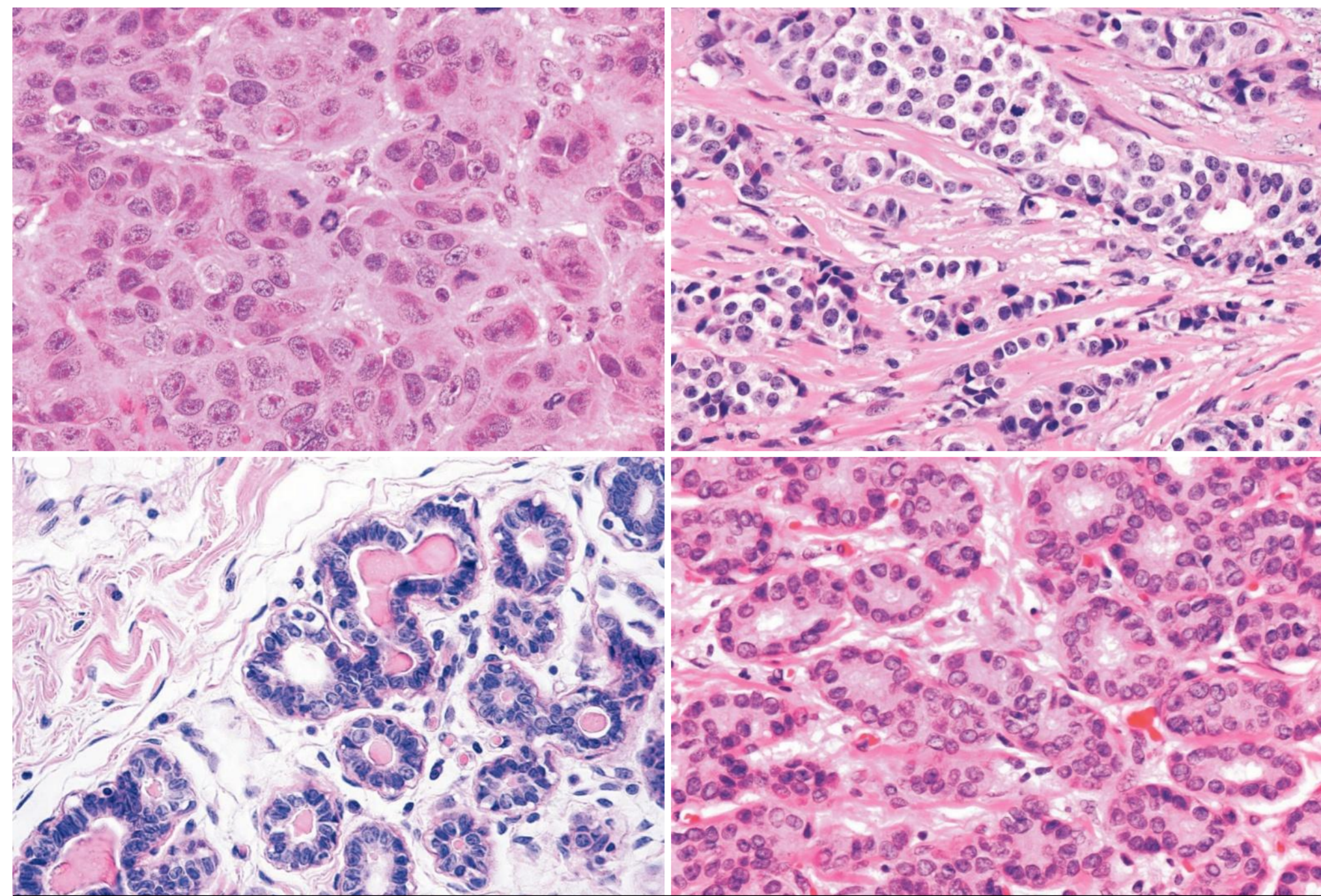
Objectiever kijken

„Dit is de toekomst!" zegt Vincent Smit, hoogleraar pathologie aan het Leids UMC. „Het is nog niet zover, maar deze technologie zal ons hopelijk helpen om objectiever naar weefsels te kijken."

De AI in de radiologie loopt ver voor, omdat röntgenbeelden al meer dan vijftien jaar digitaal worden bekeken. Dat is voor pathologie-afdelingen met hun archieven vol glaasjes met weefsel nog niet gangbaar, maar dat verandert nu snel. Sommige laboratoria, zoals in Hengelo, werken al volledig digitaal en steeds meer ziekenhuizen volgen dat voorbeeld. Van de UMC's zijn onder meer die in Utrecht en sinds kort Nijmegen ermee begonnen.

Bruikbaar in de klinische praktijk is het nog niet, zegt Smit. „Als je weet welke tumor het is, zoals in de hersentumorenstudie, dan ben je er nog niet. Om een goed behandelplan te maken gaat het juist om de subtypes, om de graad van de tumor, en om informatie over specifieke moleculen die de tumoren hebben. Die zijn nu alleen te zien met genetische tests die soms een week in beslag nemen. De hoop is dat AI iets kan leren onderscheiden in weefselplakjes dat wij pathologen niet zien."

Het algoritme van de Nijmeegse onderzoekers doet een stap in die richting. Hun systeem kan de verschillende gradaties van agressiviteit van prostaatkanker onderscheiden op plakjes weefsel. „Veel bedrijven maken algoritmes die kunnen herkennen of er sprake is van prostaatkan-

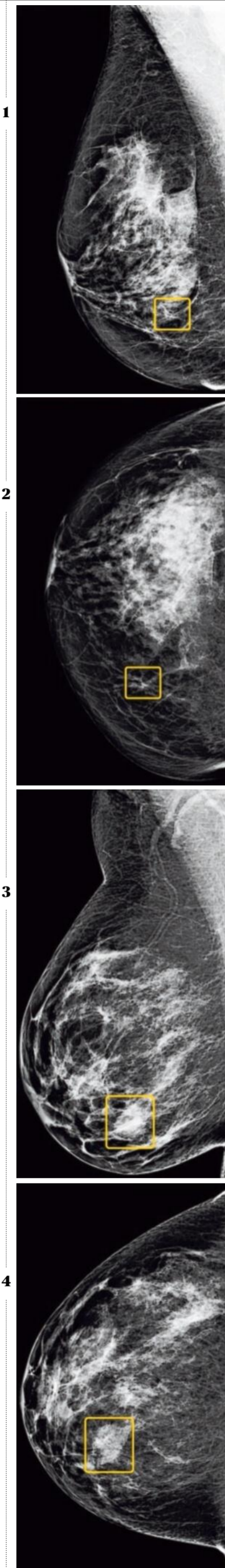


Links: Microscopisch beeld van drie stadia van borstkanker, van linksboven met de klok mee hoog-, midden- en laaggradig invasief ductaal carcinoom. Linksonder: gezond weefsel.

FOTO ANP

Bij de cover: Door de computer voorspelde agressiviteit van tumorweefsel in plakjes prostaataweefsel. Groen is gezond, geel-oranje-rood zijn toenemende gradaties van kanker.

BEELD RADBOUDUMC



ker of niet", zegt promovendus Wouter Bulten, eerste auteur van de studie. „Maar graderen is veel lastiger. Er zijn vijf zogenaemde Gleason grades. Daarmee bepaalt een arts of een man geopereerd en behandeld moet worden, of af kan wachten."

Het systeem werd getraind met miljoenen beeldjes van zesduizend biopoten met de bijbehorende gradering. „AI leert in een aantal weken waar een patholoog een heel leven over doet", zegt Bulten. In een test stelden vijftien pathologen de diagnose van honderd patiënten. „Het algoritme deed het beter dan tien van hen," aldus Bulten.

Niets dan voordelen lijkt AI te bieden: snel, consistent, onvermoeibaar, goedkoop, en een uitkomst in landen waar te weinig pathologen zijn. Maar nadelen zijn er natuurlijk ook. Een veelgehoord bezwaar is dat algoritmes een 'black box' vormen: als je zelflerende algoritmes zelf laat bepalen op welke aspecten ze moeten letten, dan weet je als gebruiker niet meer wat er gebeurt. „Onzin", vindt Bram van Ginneken, die ook meewerkte aan het artikel over prostaatkanker. „Je kunt zo'n systeem je gewoon laten vertellen waar het op let." In Nijmegen splitsen ze hiervoor de taken van de computer op. In de eerste stap leren ze de computer prostaatkanker herkennen, in een tweede stap de mate van agressiviteit. Zo kan een patholoog elke stap controleren.

Over één ding zijn alle onderzoekers het eens: het zal nog jaren duren voor de computer echt de patholoog met raad en daad terzijde kan staan. Pathologielabs die nu beginnen met digitaal werken, lopen tegen de beperkte rekenkracht van computers aan, en mensen moeten eraan wennen. Daarnaast moet nu getest worden of de bestaande algoritmes ook accurate diagnoses stellen in grotere groepen mensen, en in andere landen. De onderzoekers in Nijmegen maakten hun algoritme en hun data publiekelijk beschikbaar en schreven een wedstrijd uit voor programmeurs om het nog beter te maken.

„Het belangrijkste", zegt borstkankerpatholoog Jelle Wesseling terwijl hij zijn glaasjes opbergt, „is dat het uiteindelijk leidt tot zinnige zorg. Dat moeten we niet uit het oog verliezen. Is de patiënt uiteindelijk beter af? Voorkomen we onnodige behandelingen, vermijden we gemiste diagnoses?"

AI zal misschien preciezer kunnen beoordelen met welk type kanker je te maken hebt en welke gradering het is. Maar de behandelingen zijn nog niet zo specifiek dat je daar dan vervolgens gericht actie op kunt ondernemen, zegt Wesseling. „Neem tamoxifen, een middel dat ingezet wordt tegen hormoongevoelige borstkanker. Negen van de tien patiënten heeft geen baat bij die behandeling, maar je weet van tevoren niet wie wel en wie niet. De hoop is dat AI dit wel zal kunnen, maar dat lijkt me optimistisch."

Hij hoopt het wel, natuurlijk. „Het voorstadium van borstkanker, DCIS (ductaal carcinoom in situ), wordt dankzij het bevolkingsonderzoek bij 2.500 vrouwen per jaar geconstateerd", zegt hij. „Daar gaan vrouwen niet dood aan, maar bij 3 op de 10 ontwikkelt DCIS zich tot borstkanker. Bij wie? Als AI die eruit kan pikken, dan zou dat echt iets toevoegen."

Vier mammogrammen van twee borsten met kwaadaardige afwijkingen uit het onderzoek in Nature.

1 en 2: zij- en bovenaanzicht van een borst met kwaadaardige kalkspatjes, die alle zes de radiologen misten maar de AI correct oppikte.

3 en 4: zij- en bovenaanzicht van een borst met een kwaadaardige tumor die zes radiologen correct identificeerden maar die de AI gemist had.

BEELD NATURE