

ENDOSCOPISTEN MOGEN NIET BLIND VERTROUWEN OP DE DETECTIE DOOR COMPUTER

Artificiële intelligentie is extra gereedschap voor endoscopie

Artificiële intelligentie (AI) gaat in de toekomst een belangrijke, aanvullende rol spelen in gastro-intestinale endoscopie. Zowel bij de detectie van afwijkingen als bij de kwaliteitscontrole.

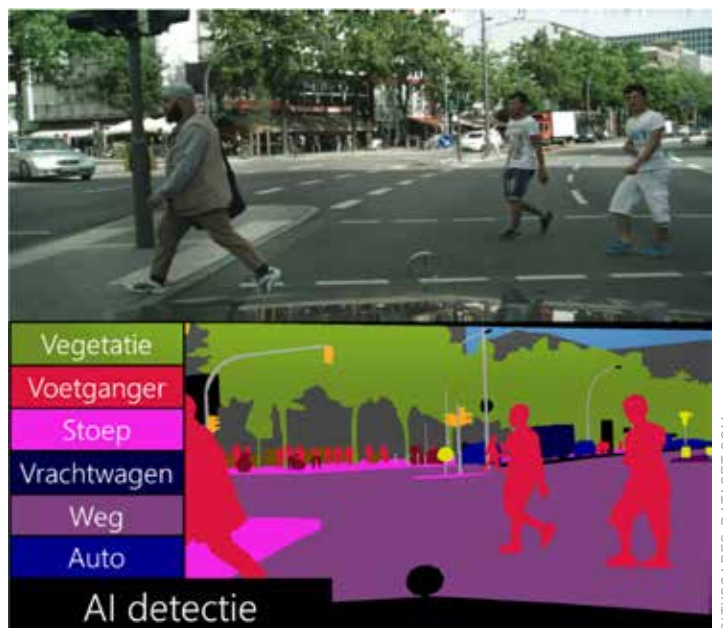


In een krap decennium heeft artificial intelligence (AI) zich ontwikkeld van een wetenschappelijke curiositeit tot een bewezen techniek die steeds verder doordringt in ons dagelijks leven. Je kunt tegenwoordig geen krant meer openslaan zonder iets over AI te lezen, in positieve of negatieve zin. En zonder dat je er erg in hebt, helpt AI je in het kiezen van de juiste Netflix-serie en het categoriseren van de foto's op je telefoon. In de nabije toekomst zal AI ons ook bij complexere taken gaan ondersteunen, zoals het besturen van onze auto (zie *figuur 1*) en het adviseren bij medische beslissingen. Waar de een zulke ontwikkelingen toejuicht, ziet de ander hierin een naderend onheil. Dit was niet anders toen de stoomtrein zijn intrede deed. Naast enthousiaste reizigers die met een snelheid van 40 km/u door de weilanden schoten, waren er kritische boeren die vreesden voor een afname van de melkproductie van hun koeien.

De technieken die ten grondslag liggen aan het recente succes van AI, stammen uit het midden van de vorige eeuw. Destijds experimenteerden wetenschappers met netwerken van kunstmatige neuronen, geïnspireerd door de neuronen in het menselijk brein. Het duurde echter nog een aantal decennia voordat de benodigde technieken werden ontwikkeld om dit soort *artificial neural networks* efficiënt voor een taak te trainen. Vervolgens was er nog een decennium nodig totdat er voldoende rekenkracht beschikbaar was dankzij de sterke opmars van computerchips, om diepere neurale netwerken (vandaar *deep learning*) te trainen die de beoogde taken, bijvoorbeeld het herkennen van handgeschreven cijfers, met menselijke nauwkeurigheid te kunnen uitvoeren.

We kunnen steeds meer aspecten van het menselijk lichaam in beeld brengen

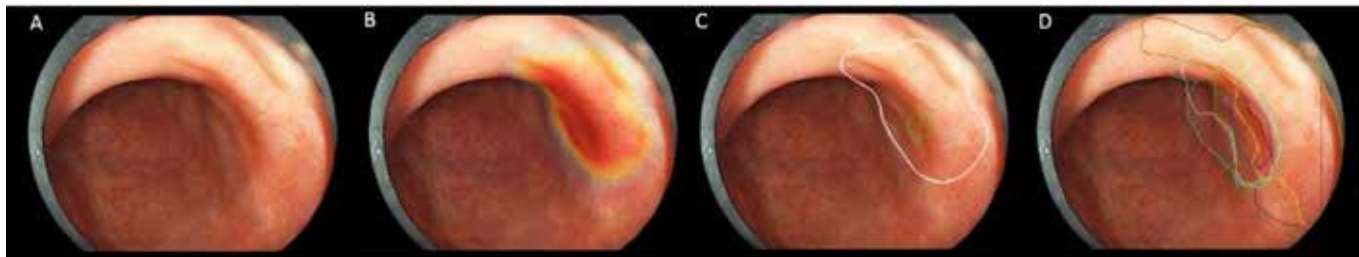
Dit uitte zich voor het eerst op grotere schaal toen in 2012 een AI-competitie voor het classificeren van beelden met ruime afstand werd gewonnen door een neuraal netwerk.¹ De gebruikte dataset, genaamd ImageNet, bevatte ruim een miljoen beelden van duizend verschillende categorieën, zoals 'boom', 'gereedschap' en 'tractor', maar ook complexere objecten zoals 'boorplatform', 'stethoscoop' en 120 verschillende soorten hondenrassen. Het neuraal netwerk bleek deze categorieën veel beter te kunnen onderscheiden dan de mens en deed dat ook nog eens meer dan duizend keer sneller.



1. AI-systeem voor zelfrijdende auto's.

Objectiever dan de mens

Technieken voor medische beeldvorming hebben ongeveer in dezelfde periode een enorme ontwikkeling doorgemaakt. We kunnen steeds meer aspecten van het menselijk lichaam in een toenemende nauwkeurigheid in beeld brengen. Binnen de endoscopie, een belangrijk aandachtsgebied van de maag-, darm- en leverarts, was deze ontwikkeling vooral gericht op het produceren van mooiere beelden (scherper, beter contrast, minder ruis) om zo endoscopisten van meer informatie te voorzien om de diagnostiek en behandeling op te baseren. Inmiddels zijn we op het punt gekomen waarop niet de endoscopische beeldresolutie, maar het menselijk oog de beperkende factor is. Een verdere toename in beeldkwaliteit kunnen mensen in steeds mindere mate waarnemen. Daarnaast verschijnen er sensoren, zoals hyperspectrale camera's, die andere licht-weefselinteracties kunnen onderscheiden dan het menselijk oog kan waarnemen. Voor de analyse van endoscopische beelden leent AI zich uitstekend. Ten eerste omdat algoritmen niet worden gelimiteerd in wat voor verschillen ze kunnen waarnemen; ze kunnen subtiele patronen herkennen die het menselijk oog ontgaan. Ten tweede omdat algoritmen kunnen leren van miljoenen voorbeelden, waardoor ze verbanden kunnen leggen die pas duidelijk worden bij een dergelijk grote sample size. Tot slot hebben algoritmen geen last van vermoeidheid of andere factoren die het beoordelingsvermogen beïnvloeden en komen ze vandaag, morgen en volgend jaar, tot dezelfde conclusie. Ze zijn wat dat betreft dus veel objectiever dan een mens. Dit gezegd hebbende, kenmerkt zich de



2. A: Een endoscopische afbeelding van een barrettblaesie; B: een AI-voorspelling in de vorm van een heat map; C: een delineatie gegenereerd door het AI-systeem met een suggestie voor een biopsie locatie; D: delineaties van zes experts op het gebied van barrettkanker.

huidige generatie AI zich vooral door een hoge nauwkeurigheid op een smal applicatiegebied. De vrees die sommigen hebben dat AI de arts zal vervangen, is dan ook ongegrond en de techniek moet gezien worden als een extra gereedschap dat de medisch specialist tot zijn of haar beschikking heeft om de kwaliteit van de zorg verder te verbeteren.

Een sprekend voorbeeld hiervan is de toepassing van AI in de detectie van vroege kanker in de barrettslokdarm. Een barrettslokdarm ontstaat als het onderste deel van de slokdarm voor langere tijd wordt blootgesteld aan terugstromend maagzuur. Hierdoor ontstaat er bij sommige patiënten een ander type slijmvlies dat beter bestand is tegen het zuur. Patiënten met een dergelijke barrettslokdarm hebben een verhoogd risico op het ontwikkelen van slokdarmkanker en ondergaan om deze reden periodieke endoscopische inspecties, waarbij het weefsel wordt bekeken en er random bipten worden afgenomen. Waar vroege vormen van slokdarmkanker nog endoscopisch te behandelen zijn met een uitstekende prognose, is er voor gevorderde tumoren een ingrijpende operatie nodig en is de vijfjaarsoverleving minder dan 50 procent. De visuele kenmerken van vroege barretneoplasie zijn echter erg subtiel en worden regelmatig niet als zodanig herkend, mede ook door de relatief lage incidentie van deze afwijkingen: een mdl-arts ziet gemiddeld minder dan één barrettvroegcarcinoom per jaar.

Valkuilen

Met het doel om endoscopisten te ondersteunen zijn we bijna tien jaar geleden begonnen met het ontwikkelen van een systeem dat vroege kanker in de barrettslokdarm kan herkennen. Recentelijk hebben we de laatste resultaten van ons *deep learning* AI-systeem gepresenteerd (zie *figuur 2*), waarin ons systeem een betere nauwkeurigheid liet zien dan 53 endoscopisten uit verschillende Europese ziekenhuizen.² Vervolgens hebben we dit systeem getest op twintig patiënten in een pilotstudie in de kliniek, waarbij het AI-systeem de bioptlocatie moest aangeven in het beeld.³ Deze studie

liet een vergelijkbare nauwkeurigheid zien en heeft ons veel geleerd over de interactie tussen het endoscopieteam en het AI-systeem. Het hele ontwikkelingsproces heeft ons daarbij inzicht gegeven in de translatie van AI naar de praktijk en de mogelijke valkuilen.

Het algoritme moet leren waar de afwijking zich bevindt

Het toepassen van AI in het medische domein kent een aantal extra uitdagingen ten opzichte van andere meer generieke AI-toepassingen, zoals bijvoorbeeld zelfrijdende auto's. Allereerst is de hoeveelheid data die beschikbaar is om het systeem mee te trainen erg beperkt. Waar het eerdergenoemde voorbeeld van ImageNet ruim een miljoen beelden bevat, zijn er voor medische toepassingen vaak maar hooguit een aantal duizenden voorbeelden beschikbaar. Daarnaast volstaat het in veel gevallen niet om als gouden standaard de beelden te laten annoteren ('dit is een kat, dit is een boorplatform') maar moet het algoritme ook leren waar in het beeld de afwijking zich bevindt en wat de histologische correlatie is. Ook in termen van variabiliteit van de gouden standaard, is er tussen annotaties van honden en boorplatforms in beelden een stuk minder variatie te verwachten dan voor annotaties van neoplastische laesies. Ten slotte zijn er nogal wat valkuilen zoals vormen van bias die kunnen worden geïntroduceerd tijdens de verzameling van data of tijdens de verdeling van data in training- en testsets.

Getraind netwerk

Deze problemen vragen daarom om specifieke strategieën in het ontwikkelen van AI-algoritmes in de endoscopie. Zo is het bijvoorbeeld mogelijk om de kennis van een neurale netwerk – ontwikkeld op breed beschikbare, niet-medische data (zoals dagelijkse gebruiksvoorwerpen en dieren uit ImageNet) – over te dragen naar een nieuw neurale netwerk voor een andere toepassing, bijvoorbeeld voor het detecteren van colonpoliepen, waarvoor er aanzienlijk minder data beschikbaar is. Deze techniek wordt *transfer learning* genoemd en ligt ten grondslag aan vrijwel alle succesvolle AI-applicaties binnen het medisch domein.⁴ Hierbij wordt het netwerk dus eerst getraind op een grote set, en vervolgens gefinetuned met de daadwerkelijke data van het op te lossen AI-probleem.

Het succes van deze techniek kan worden verklaard door de manier waarop neurale netwerken informatie uit een beeld halen ('training'): de eerste lagen extraheren simpele eigenschappen zoals randen en kleuren, deze worden vervolgens in de daaropvolgende lagen gecombineerd tot vormen, objecten en ten slotte tot een classificatie. Omdat de simpele eigenschappen uit de eerste lagen zowel nuttig zijn voor de classificatie van katten en honden als voor veel medische beelden, kunnen deze lagen worden hergebruikt, en hoeven ze niet opnieuw te worden getraind. Omdat er slechts een deel van het netwerk getraind hoeft te worden, zijn daar minder data voor nodig. Dit gebeurt echter dus vaak met datasets die weinig gelijkenis vertonen met het uiteindelijke AI-probleem waar het systeem voor ontwikkeld wordt. Logischerwijs zou *transfer learning* met beelden die meer gelijkenis vertonen met de uiteindelijke input voor het AI-systeem tot een beter resultaat moeten leiden. In de eerste trainingsfase wordt in nagenoeg alle gevallen met ImageNet (afbeeldingen van gebruiksartikelen, voertuigen, dieren enzovoorts) gewerkt. Onlangs hebben wij aangetoond dat *transfer learning*, waarbij voor de eerste trainingsfase gebruik wordt gemaakt van een grote dataset met generieke endoscopische beelden (GastroNet), resulteert in betere AI-prestaties voor endoscopische applicaties.²

Wisselwerking

Wat zijn de toekomstige applicaties voor AI binnen de endoscopie? Naast het gebruik van AI voor het detecteren van afwijkingen (zoals poliepen of neoplastische laesies), zijn er ook mogelijkheden voor het gebruik van AI voor kwaliteitscontrole. Een algoritme dat aangeeft dat er een deel van het colon onvoldoende zichtbaar is gemaakt, is mogelijk waardevoller dan een algoritme dat endoscopisten helpt beter poliepen te detecteren in dat deel van het colon dat de endoscopist daadwerkelijk in beeld brengt. Dergelijke feedback op de endoscopische kwaliteit is ook van belang om een belangrijke potentiële valkuil in de toepassing van AI te voorkomen: het gevaar bestaat namelijk dat endosco-



MERLIN DALEMAN, HH

pisten blind gaan vertrouwen op de endoscopische detectie door het algoritme. Het succes van elk AI-systeem staat of valt namelijk met de kwaliteit van de input: wat niet (of niet goed) in beeld wordt gebracht kan niet door een algoritme worden beoordeeld. Zowel voor de input als voor de interpretatie van de output van het algoritme is dus een wisselwerking tussen de endoscopist en het algoritme van belang. ■

contact

j.j.bergman@amsterdamumc.nl
cc: redactie@medischcontact.nl

web

Literatuurverwijzingen vindt u onder dit artikel op medischcontact.nl/ artikelen.